明 細 書

セッション中継装置及び中継方法

技術分野

[0001] 本発明はセッション中継装置及びそれに用いるセッション中継方法に関し、特にTCP(Transmission Control Protocol)セッション間でデータの中継を行う装置に関する。

背景技術

[0002] 一般的に、通信アプリケーションにおいては、送信端末と受信端末との間で通信セッションを確立し、確立したセッション上で通信を行う。しかしながら、送信端末と受信端末との間の伝播遅延時間が非常に長い場合、もしくは有線と無線とのように特性の異なるネットワークを横断して通信する場合には、送信端末と受信端末との間の通信のスループットが低下する。

[0003] この問題を解決するための方法としては、下記の文献1ー3に開示されているような方式が存在する。この方式では、送信端末と受信端末との間を一つのセッションで通信を行うのではなく、送信端末と受信端末との間に中継装置を設置する。そして、送信端末から中継装置へのセッションと中継装置から受信端末へのセッションの2つのセッション間でデータの中継を行うことによって通信が行われる。

[0004] 文献1:特開平11−252179号公報

文献2:特開2002−281104号公報

文献3:jay Bakre and B. R. Badrinath, "I−TCP;Indirect TCP for Mobile Host", Department of Computer Science Rutgers University, DSC−TR−314, 1994(http://www. it. iitb. ac. in/it644/papers/i−tcp. pdf)

中継装置においては、伝播遅延時間が非常に長い場合でもスループットを低下させないために、大きな送信バッファを持つことが有効であることが知られている。しかしながら、大きな送信バッファを持った場合、特にTCPのスロースタート(Slow Start)時において、パケットがバースト的に出力されるためにネットワークの輻輳を発生さ

せ、結果としてスループットの低下を招く場合がある。

[0005] このような問題を解決する方法としては、下記の文献4に開示されているような、TCPセッションからバースト的にパケットを出力しないように出力するパケットの間隔を調整する方法がある。

[0006] 文献4:A. Aagarwal, S. Savage, and T. Anderson. "Understanding the Performance of TCP Pacing", in proceedings of IEEE INFOCOM´2000

また、他の方法としては、TCPセッションからの出力パケットをキューイングしてスケジューラで出力制御を行うことによって、任意の帯域でのパケット出力が可能となり、他のTCPセッションのスループットを制限し、特定のセッションのスループットを向上させることも可能である。

[0007] セッションの中継を行う場合、最も問題となることは、一般的にTCPセッションの処理において処理負荷が高く、高速な中継処理が難しいという問題がある。また、TCPセッションからのパケット出力間隔を調整する場合には、処理負荷がさらに高くなる。

[0008] 上記のTCPセッションの処理を高速化する方式としては、以下に示すような方式が存在する。第1の方式としては、OS(Operation System)のカーネルとアプリケーションプログラムとの間のデータ転送にゼロコピー方式を用いることによって、中継装置の処理負荷を下げる方式がある。この第1の方式では、実際にデータの送受信処理を行うカーネルプログラムと、データの中継処理を行うアプリケーションプログラムとの間で、データの受け渡しを行う際、物理的なデータのコピーを行わず、ページマッピングによる仮想的なデータの移動を行う。

[0009] 第2の方式としては、中継処理を行う際、再送制御のみを行う方式がある。この第2の方式では、中継装置でセッションの終端を行わず、送受信端末間で1本のセッションを設定する。そして、中継装置においては送信端末から受信したパケットを受信端末へと出力する際、単にパケットを転送するだけでなく、パケットを中継装置内に保存しておく。

[0010] 中継装置では受信端末から送信端末に返されるACK(acknowledgement)パケットを監視し、もし中継装置と受信端末との間でパケット廃棄を検出すると、保存して

おいたパケットを受信端末に出力することで、送信端末からのパケットの再送を防ぎ、パケット廃棄による送信端末と受信端末との間のスループットの低下を防いでいる。

[0011] この第2の方式では、セッション間の中継処理を行わず、再送処理を行うのみであり、中継処理を行う場合に比べて中継装置の負荷を下げることができる。

[0012] 第3の方式としては、中継の必要が無いセッションの中継処理を行わず、中継を行う必要があるセッションのみ中継処理を行うことで、中継装置の処理負荷を下げる方式がある。例えば、この方式としては、下記の文献5に開示されているような、セッションのスループットを計測し、中継処理を行うことで、スループットが向上するセッションのみを選択して中継処理を行う方式がある。

[0013] 文献5:特開平11−112576号公報

また、上記の方式としては、下記の文献6に開示されているように、HTTP(Hyper Text Transfer Protocol)のリクエストを送信する際に中継処理を行わず、データ転送を行う際のみ中継処理を行う方式がある。

[0014] 文献6:特開2002−312261号公報

上述した従来のセッション中継方式では、第1の方式の場合、データコピー以外の処理負荷も高いという問題がある。TCP pacing方式やスケジューラによる出力制御を行う場合、TCPの処理負荷に加えて、パケット出力制御の負荷が加わるため、全体の処理負荷が高くなる。

[0015] また、TCPセッションの中継だけではなく、例えばiSCSI(internet Small Computer System Interface)のような上位レイヤプロトコルの中継処理を行う場合、さらに上位レイヤでの輻輳制御に基づくパケット出力制御が加わり、さらに処理負荷が高くなる。特に、中継するセッション数が多い場合には、セッション間でパケットの出力制御を行う処理の負荷が高くなる。

[0016] さらに、従来のセッション中継方式では、第1の方式の場合、パケット長に制限があるという問題がある。TCPレイヤとアプリケーションとの間でページマッピングによる仮想的なデータの移動を行うためには、CPU(中央処理装置)のページサイズとパケット長とが一致している必要がある。

[0017] 一般的に、セッション中継装置は送受信端末とは独立して存在し、送信端末及び

受信端末が用いるパケット長を予め想定しておくことはできないため、帯域制御装置のページサイズと異なるパケット長を用いる送信端末及び受信端末に対しては、ゼロコピーによる処理負荷低減が期待できない。

[0018] 一方、従来のセッション中継方式では、第2の方式の場合、パケット廃棄以外の要因によるスループット低下を改善することができないという問題がある。セッションの中継処理を行う場合には、セッション中継装置では送信端末からのパケット受信に対してすぐさま送信端末に対して受信確認のためのACKパケットを返送するが、この第2の方式では受信端末が送信端末に対してACKパケットを返すのみであり、セッション中継装置自身はACKパケットを返さない。そのため、伝播遅延が大きな環境では第2の方式によるスループット向上の効果が小さい。

[0019] また、従来のセッション中継方式では、第3の方式の場合、中継する必要のあるセッション数が多い場合に処理負荷が軽減されないという問題がある。

発明の開示

[0020] そこで、本発明の目的は上記の問題点を解消し、中継するセッション数が多く、パケットの出力制御を行う場合でも、高速にセッション間の中継処理を行うことができるセッション中継装置及びそれに用いるセッション中継方法を提供することにある。

[0021] 本発明によるセッション中継装置は、複数のレイヤにおける輻輳制御処理とパケット出力制御処理とを含むセッション中継処理を行うセッション中継装置であって、前記複数のレイヤ各々が前記輻輳制御情報の作成のみを行い、前記パケット出力制御処理をIP(Internet Protocol)レイヤのスケジューラに集約している。

[0022] 本発明による他のセッション中継装置は、送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記受信端末と前記受信端末との間の通信を実現するセッション中継装置であって、

前記送信端末に向けたセッションからのデータを受信する受信セッション処理手段と、前記受信端末に向けたセッションへとデータを送信する送信セッション処理手段と、前記送信端末へと出力するデータを一時蓄えておく送信バッファと、前記送信バッファからのパケット出力を制御するパケットスケジューラと、前記パケットスケジューラの制御に応答して前記送信バッファに蓄えられたデータを出力制御する出力制御手

段とを備え、

前記送信セッション処理手段において当該レイヤで出力が許可されているデータ量を計算し、これに基づいて前記パケットスケジューラが前記パケット出力を制御している。

[0023] 本発明による別のセッション中継装置は、送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記送信端末と前記受信端末との間の通信を実現するセッション中継装置であって、

複数のレイヤに対応して設けられかつ前記送信端末に向けたセッションからのデータを受信する受信セッション処理手段と、前記複数のレイヤに対応して設けられかつ前記受信端末に向けたセッションへとデータを送信する送信セッション処理手段と、前記送信端末へと出力するデータを一時蓄えておく送信バッファと、前記送信バッファからのパケット出力を制御するパケットスケジューラとを備え、

前記送信セッション処理手段各々において当該レイヤで出力が許可されているデータ量を計算し、前記複数のレイヤ全てで共通に許可されるデータ量に基づいて前記パケットスケジューラが前記パケット出力を制御している。

[0024] 本発明によるセッション中継方法は、複数のレイヤにおける輻輳制御処理とパケット出力制御処理とを含むセッション中継処理を行うセッション中継装置のセッション中継方法であって、前記複数のレイヤ各々で前記輻輳制御情報の作成のみを行い、前記パケット出力制御処理をIP(Internet Protocol)レイヤのスケジューラに集約している。

[0025] 本発明による他のセッション中継方法は、送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記受信端末と前記受信端末との間の通信を実現するセッション中継装置のセッション中継方法であって、前記セッション中継装置側に、前記送信端末に向けたセッションからのデータを受信する受信セッション処理と、前記受信端末に向けたセッションへとデータを送信する送信セッション処理と、前記送信端末へと出力するデータを送信バッファに一時蓄えておく処理と、前記送信バッファからのパケット出力をパケットスケジューラにて制御する処理と、前記パケットスケジューラの制御に応答して前記送信バッファに蓄えられたデ

ータを出力制御手段にて出力制御する処理とを備え、前記送信セッション処理において当該レイヤで出力が許可されているデータ量を計算し、これに基づいて前記パケットスケジューラが前記パケット出力を制御している。

[0026] 本発明による別のセッション中継方法は、送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記送信端末と前記受信端末との間の通信を実現するセッション中継装置のセッション中継方法であって、前記セッション中継装置側に、複数のレイヤ各々において前記送信端末に向けたセッションからのデータを受信する受信セッション処理と、前記複数のレイヤ各々において前記受信端末に向けたセッションへとデータを送信する送信セッション処理と、前記送信端末へと出力するデータを送信バッファに一時蓄えておく処理と、前記送信バッファからのパケット出力をパケットスケジューラにて制御する処理とを備え、前記送信セッション処理各々において当該レイヤで出力が許可されているデータ量を計算し、前記複数のレイヤ全てで共通に許可されるデータ量に基づいて前記パケットスケジューラが前記パケット出力を制御している。

[0027] すなわち、本発明のセッション中継装置は、TCP(Transmission Control Protocol)レイヤにおいてパケット出力の制御を行わず、TCPレイヤでパケット出力のための制御情報を生成するのみとし、IP(Internet Protocol)レイヤにおいてパケットスケジューラを用いてパケット出力の制御を行う。

[0028] また、本発明のセッション中継装置では、iSCSI(internet Small Computer System Interface)のような上位レイヤプロトコルにおける輻輳制御に関しても、制御情報の作成のみを上位レイヤで行い、実際のパケット出力制御にはIPレイヤのパケットスケジューラを用いる。

[0029] これによって、本発明のセッション中継装置では、複数のレイヤにおけるデータ出力制御を統合することができるため、パケット出力に関する処理負荷を軽減することが可能となる。

[0030] さらに、本発明のセッション中継装置では、アプリケーションへのデータコピーを行わないため、ページマッピングによるデータの移動の必要がない。すなわち、受信パケットはTCPレイヤやその他の上位レイヤ、アプリケーションの受信バッファに格納さ

れることなく、またTCPレイヤやその他の上位レイヤ、アプリケーションの送信バッファに格納に格納されることもなく、直接IPレイヤの送信バッファに格納される。そのため、本発明のセッション中継装置では、セッション中継装置内のデータの移動は発生せず、ページマッピングによるデータの移動の必要がない。

[0031] さらにまた、本発明のセッション中継装置では、中継処理をパケットの再送制御にのみを行うのではなく、セッションを一旦終端して完全な中継処理を行っているため、パケット廃棄以外の要因によるスループット低下を改善することが可能となる。

[0032] 本発明のセッション中継装置では、セッション間の中継処理を高速化することで、中継セッション数が多い場合でも高速に処理を行うことが可能となる。

図面の簡単な説明

[0033] [図1]図1は、本発明の第1の実施形態によるセッション中継装置を含む伝送システムの構成を示すブロック図である。

[図2]図2は、本発明の第1の実施形態によるセッション中継装置の構成を示すブロック図である。

[図3]図3は、図2のパケットスケジューラの構成を示すブロック図である。

[図4]図4は、本発明の第1の実施形態によるセッション中継装置の動作を示すフローチャートである。

[図5]図5は、本発明の第1の実施形態によるセッション中継装置の動作を示すフローチャートである。

[図6]図6は、図3に示すパケットスケジューラの動作を示すフローチャートである。

[図7]図7は、送信待ちバイト数を示す模式図である。

[図8]図8は、本発明の第2の実施形態によるパケットスケジューラの構成を示すブロック図である。

[図9]図9は、本発明の第3の実施形態によるセッション中継装置の構成を示すブロック図である。

[図10]図10は、本発明の第4の実施形態によるセッション中継装置の構成を示すブロック図である。

[図11]図11は、本発明の第4の実施形態における送信待ちバイト数を説明する模式

図である。

[図12]図12は、本発明の第5の実施形態によるセッション中継装置の構成を示すブロック図である。

[図13]図13は、本発明の第5の実施形態によるセッション中継装置(送信端末)及びセッション中継装置(受信端末)の間のデータの流れを示すブロック図である。

発明を実施するための最良な形態

[0034] 次に、本発明の実施の形態について図面を参照して説明する。

(第1の実施形態)

図1は本発明の第1の実施形態によるセッション中継装置1を含む伝送システムの構成を示すブロック図である。図1において、本実施形態によるセッション中継装置1はセッション識別部11と、セッション中継部12−1ー12−Nと、出力制御部14とから構成され、受信端末2及び送信端末3に接続されている。

[0035] まず、送信端末3から受信端末2へデータを送る場合、送信端末3からのデータパケットはセッション中継部12−1の受信セッション処理部(図示せず)で処理され、その結果、ACK(acknowledgement)パケットが送信端末3へと返信される。

[0036] セッション中継部12−1の受信セッション処理部で受取られたデータはセッション中継部12−1の送信セッション処理部(図示せず)へと送られ、ここから受信端末2へとデータパケットが送信される。これに対して、受信端末2が返信したACKパケットはセッション中継部12−1の送信セッション処理部で処理される。

[0037] 同様に、受信端末2から送信端末3へデータを送る場合、受信端末2からのデータパケットはセッション中継部12−2の受信セッション処理部(図示せず)で処理され、その結果、ACKパケットが受信端末2へと返信される。

[0038] セッション中継部12−2の受信セッション処理部で受取られたデータはセッション中継部12−2の送信セッション処理部(図示せず)へと送られ、ここから送信端末3へとデータパケットが送信される。これに対して、送信端末3が返信したACKパケットはセッション中継部12−2の送信セッション処理部で処理される。

[0039] 図2は本発明の第1の実施形態によるセッション中継装置の構成を示すブロック図である。図2において、セッション中継装置1はセッション識別部11と、セッション中継

部12−1ー12−Nと、パケットスケジューラ13と、出力制御部14とから構成されている。

[0040] セッション識別部11は到着したパケットが属するセッションを決定する。セッション中継部12−1ー12−Nは送信端末3とのセッションと受信端末2とのセッションとの間で中継を行う。パケットスケジューラ13は各セッション中継部12−1ー12−Nからのパケット出力を制御する。出力制御部14はパケットスケジューラ13からの指示に基づいて各セッション中継部12−1ー12−Nからのパケット出力を行う。

[0041] また、セッション中継部12−1は受信端末2へとデータを送信するセッションの処理を行う送信セッション処理部121−1と、受信したデータを送信終了まで蓄えておく送信バッファ122−1と、送信端末3からデータを受信するセッションの処理を行う受信セッション処理部123−1とから構成されている。尚、図示していないが、セッション中継部12−2ー12−Nの構成は上記のセッション中継部12−1の構成と同様である。

[0042] 図3は図2のパケットスケジューラ13の構成を示すブロック図である。図3において、パケットスケジューラ13はリスト振り分け部131と、状態更新部132と、状態変数保存部133と、リセット制御部134と、送信可能リスト135と、送信待ちリスト136とから構成されている。

[0043] 送信可能リスト135はパケット出力が可能なセッションの識別子を保持し、送信待ちリスト136は送信待ち状態にあるセッションの識別子を保持する。リスト振り分け部131はデータパケットセッションの識別子もしくはACKパケットを受信したセッションの識別子を送信可能リスト135もしくは送信待ちリスト136に振り分ける。

[0044] 状態更新部132は送信可能リスト135からセッションの識別子を取出して出力制御部14に通知し、かつ該セッションの状態を更新する。状態変数保存部133は各セッションの状態を保持し、リセット制御部134は送信待ちリスト136で管理されているセッションの識別子を送信可能リスト135へと移動させる。

[0045] TCP(Transmission Control Protocol)セッションでは、通常、送信端末3と受信端末2との間の双方向の通信を行う。そのため、本実施形態においては、一組の送信端末3及び受信端末2に対して2つのセッション中継部を使用するものとし、夫々の方向へのデータ通信に対しては夫々対応するセッション中継部を使用する。

[0046] したがって、セッション中継部12−1ー12−Nは複数の送信端末3及び受信端末2の組に対して夫々2つずつ用意され、セッション中継部12−1ー12−N各々は夫々の送信端末3からのセッションから対応する受信端末2へのセッションへと、もしくは夫々の受信端末2からのセッションから対応する送信端末3へのセッションへとデータを中継する処理を行う。

[0047] 尚、TCPセッションにおいては、ある方向のデータパケットとその反対方向へのACKパケットとが1つのパケット上に統合される場合があるが(ACKのpiggy back)、本実施形態では説明を簡単化するために、このような動作に関しての説明については省略する。

[0048] 図4及び図5は本発明の第1の実施形態によるセッション中継装置1の動作を示すフローチャートであり、図6は図3に示すパケットスケジューラ13の動作を示すフローチャートである。これら図1ー図6を参照して本発明の第1の実施形態によるセッション中継装置1の動作について説明する。

[0049] 本実施形態によるセッション中継装置1にパケットが入力された時(図4ステップS1)、セッション識別部11はパケットのヘッダを参照し、送信元IP(Internet Protocol)アドレス、送信先IPアドレス、第四層プロトコル番号、送信元第四層ポート番号、送信先第四層ポート番号等に基づいて、パケットが属するセッションを決定する(図4ステップS2)。

[0050] セッション識別部11はパケットがデータパケットであれば(図4ステップS3)、対応するセッション中継部12−1ー12−Nの受信セッション処理部123−1ー123−N(受信セッション処理部123−2ー123−Nは図示せず)へと渡す(図4ステップS4)。

[0051] また、セッション識別部11はパケットがACKパケットであれば(図4ステップS3)、対応するセッション中継部12−1ー12−Nの送信セッション処理部121−1ー121−N(送信セッション処理部121−2ー121−Nは図示せず)へと渡す(図4ステップS10)。

[0052] 受信セッション処理部123−1ー123−Nでは入力されたデータパケットをシーケンス番号順に並べ替えて送信バッファ122−1ー122−N(送信バッファ122−2ー122−Nは図示せず)へと格納する(図4ステップS5)。また、受信セッション処理部123−1ー123−Nは受信したデータパケットが正しいシーケンス番号を持つものであれば(

図4ステップS6)、すなわち連続して受信しているデータの最後尾のシーケンス番号と連続していれば、送信端末3に対してデータの受信確認及び広告ウインドサイズを通知するために、ACKパケットを出力制御部14を通して返送する(図4ステップS7)。

[0053] さらに、受信セッション処理部123−1ー123−Nは連続して受信した最後のパケットのシーケンス番号を基にACKパケット(重複ACKパケット)を返送し、送信端末3に対してパケットの未到着を通知する。

[0054] 上記の処理に関しては、TCP／IP Illustrated, Volume 1:The Protocols, Addison−Wesley, 1994, ISBN 0−201−63346−9(以下、文献7とする)に詳しく記載されているので、その説明については詳述しない。

[0055] ACKパケットは生成された後、すぐさま出力制御部14から出力されるか、もしくは対応する逆方向のセッションの送信バッファ122−1ー122−Nに格納され、逆方向のセッションのデータパケットとともに、パケットスケジューラ13の指示で出力される。

[0056] また、受信セッション処理部123−1ー123−Nでは、連続して受信しているデータの最後尾のシーケンス番号をパケットスケジューラ13へと通知する(図4ステップS8)。

[0057] 送信セッション処理部121−1ー121−Nでは入力されたACKパケットを基に輻輳ウインドサイズの変更を行い(図4ステップS9)、ACKパケットが重複ACKでなければ受信が確認されたデータを送信バッファ122−1ー122−Nから消去し(図5ステップS12, S13)、重複ACKであれば必要に応じてデータの再送処理を行う(図5ステップS12, S11)。尚、この処理に関しても上記の文献7に詳しく記載されているので、その説明については詳述しない。

[0058] 送信セッション処理部121−1ー121−Nは受信したACKパケットに記されている広告ウインドと、更新した輻輳ウインドとをパケットスケジューラ13に通知する(図5ステップS15)。また、送信セッション処理部121−1ー121−Nは再送がタイムアウトした場合(図5ステップS14)、受信確認済みの最後のシーケンス番号も通知する(図5ステップS16)。

[0059] 尚、本実施形態ではパケットスケジューラ13からパケット出力の指示があった時に

のみパケットの出力を行うため、ここでは重複ACKを受信した時にすぐさまパケットの再送を行うのではなく、次に出力すべきパケットとして再送するパケットのシーケンス番号を記憶しておくのみである。

[0060] 送信バッファ122−1ー122−Nからのパケット出力は、パケットスケジューラ13からの指示に基づいて行われる。パケットスケジューラ13からパケット出力の指示があると、出力制御部14は送信バッファ122−1ー122−Nからパケットを1つ取出して出力回線に出力し、出力したパケットのパケット長をパケットスケジューラ13に通知する。

[0061] 送信バッファ122−1ー122−Nから出力するパケットは、再送処理を行う場合に記憶しておいたシーケンス番号のパケットであり、さもなければ未送信のパケットのうち最もシーケンス番号の小さいパケットである。

[0062] 尚、該セッションがTCPセッションでなければ、受信セッション処理部123−1ー123−Nでは受信したパケットをそのまま到着順に送信バッファ122−1ー122−Nに格納するのみである。送信セッション処理部121−1ー121−NではACKパケットを受信せず、パケットスケジューラ13に対しては送信バッファ122−1ー122−Nのキュー長を広告ウインド及び輻輳ウインドとして通知することで、送信バッファ122−1ー122−Nにパケットがある限り、パケットスケジューラ13にパケット出力を要求し続ける。

[0063] パケットスケジューラ13では、状態変数保存部133において、セッション毎に、割り当てウエイト、送信可能バイト数、送信待ちバイト数の3つのパラメータを保持する。

[0064] 割り当てウエイトは該セッションに割り当てられたウエイトである。パケットスケジューラ13は一定周期、あるいはパケット送信可能なセッションがなくなる毎にリセットを行い(図6ステップS26, S27)、このリセット1周期内に送信可能なバイト数が割り当てウエイトである。

[0065] 送信可能バイト数は現時点から次のリセットまでに送信可能なバイト数であり、初期値は割り当てウエイトであって(図6ステップS21)、パケットを出力する毎にパケット長分だけ減算される(図6ステップS22, S23)。

[0066] 送信待ちバイト数はTCPセッションの場合、TCPレイヤが送信許可したシーケンス番号から既にパケットスケジューラ13が送信を行ったシーケンス番号を引いたものであり、min(連続して受信しているデータの最後尾のシーケンス番号＋1、受信端末が

示した広告ウインド、該セッションの輻輳ウインド)ー送信済みシーケンス番号となる。TCPセッションでない場合、送信待ちバイト数は送信バッファ122−1ー122−N内のキュー長である。

[0067] 図7は送信待ちバイト数を示す模式図である。図7において、パケット出力が可能なセッション、すなわち送信可能バイト数と送信待ちバイト数とがともに1あるいはMSS(Maximum Segment Size)以上のセッションの識別子は送信可能リスト135で管理され(図6ステップS24, S25)、送信可能バイト数がリセットされた後にパケット出力が可能となるセッション、すなわち送信待ちバイト数が1あるいはMSS以上あるが、送信可能バイト数が1あるいはMSS未満であるセッションの識別子は送信待ちリスト136で管理される(図6ステップS27ーS29)。

[0068] 次に、図3を参照してパケットスケジューラ13の動作について説明する。リスト振り分け部131はセッション中継部12−1ー12−Nから、送信端末3から連続して受信している最後尾のシーケンス番号や、受信端末2から受信した広告ウインド、更新した輻輳ウインドを受取ると、図7に示すようにして送信待ちバイト数を更新する。

[0069] また、リスト振り分け部131は出力制御部14において再送タイマがタイムアウトした際にも、送信済みシーケンス番号を受信確認済みのシーケンス番号まで戻し、送信待ちバイト数を更新する。リスト振り分け部131は更新前の送信待ちバイト数が1あるいはMSS未満であれば、該セッションの識別子はまだ送信可能リスト135あるいは送信待ちリスト136で管理されていないため、更新後の送信待ちバイト数及び送信可能バイト数を基に該セッションの識別子を送信可能リスト135あるいは送信待ちリスト136に新たに格納する。

[0070] 状態更新部132は送信可能リスト135の先頭から送信可能なセッションの識別子を1つ取出し、これを出力制御部14に通知してパケット出力を行わせる。その後、状態更新部132は送信済みシーケンス番号に出力したパケット長を加算した後送信待ちバイト数を更新し、送信可能バイト数から出力したパケット長を減算する。

[0071] 状態更新部132は送信待ちバイト数及び送信可能バイト数にしたがって、改めて該セッションの識別子を送信可能リスト135あるいは送信待ちリスト136に格納する。すなわち、状態更新部132は送信待ちバイト数及び送信可能バイト数がともに1あるい

はMSS以上であれば送信可能リスト135に格納し、送信待ちバイト数が1あるいはMSS以上であるが送信可能バイト数が1あるいはMSS未満であれば送信待ちリスト136に格納する。

[0072] 状態更新部132は送信可能リスト135が空になれば、全てのセッションの送信可能バイト数をリセット、すなわち送信可能バイト数に割り当てウエイトを加え、もし送信可能バイト数が割り当てウエイト以上となれば送信可能バイト数を割り当てウエイトの値とする[送信可能バイト数＝min(送信可能バイト数＋割り当てウエイト, 割り当てウエイト)]。

[0073] この処理によって、送信待ち状態にあったセッションは全て送信可能な状態に変化するため、送信待ちリスト136で管理されていたセッションを全て送信可能リスト135へと移動する。

[0074] このリセット処理を行う前に、もし送信可能バイト数が1あるいはMSS未満ではなく、かつ該セッションの送信バッファ122−1ー122−N内に送信待ちのデータがあれば、該セッションからの出力帯域に余裕はあるが、送信セッション処理部121−1ー121−NにおいてTCP制御によって送信を止められている状態であるとする。

[0075] この場合には、将来の送信再開に備えて帯域を蓄えておくため、送信可能バイト数と割り当てウエイトとの和が割り当てウエイトを超えた場合でも、ある一定値までは送信可能バイト数を蓄えておく[送信可能バイト数＝min(送信可能バイト数＋割り当てウエイト, 送信可能バイト数上限値)]。

[0076] 上述したように、本実施形態では、中継するパケットが受信時の送信バッファ122−1ー122−Nへの格納と、送信時の送信バッファ122−1ー122−Nからの取出しの際にのみ移動するため、データ移動のオーバヘッドが小さい。また、TCPレイヤからの情報を用いたパケットスケジューラ13によってパケット出力制御を行うため、送信セッション処理部121−1ー121−Nが直接パケット出力するよりも、少ない処理負荷で帯域制御を行いながらのパケット出力を行うことができる。

(第2の実施形態)

本発明の第2の実施形態によるセッション中継装置の構成は図2に示す本発明の第1の実施形態によるセッション中継装置1の構成と同様である。

[0077] 図8は本発明の第2の実施形態によるパケットスケジューラの構成を示すブロック図である。図8において、本発明の第2の実施形態によるパケットスケジューラ16は割り当てウエイト変更部161及び制御パラメータ変更部162を加えた以外は図3に示す本発明の第1の実施形態と同様の構成となっており、同一構成要素には同一符号を付してある。また、同一構成要素の動作は本発明の第1の実施形態と同様である。

[0078] この図8を参照して本発明の第2の実施形態によるパケットスケジューラ16の動作について説明する。ここでは、本発明の第1の実施形態によるパケットスケジューラ13との相違点のみを説明する。

[0079] 本実施形態によるセッション中継装置では、パケットスケジューラ16内の制御パラメータ変更部162が割り当てウエイトとして設定した帯域にしたがって該セッションのTCP制御パラメータの値を動的に変更している。すなわち、制御パラメータ変更部162は該セッションの送信可能バイト数が割り当てウエイトよりも大きい予め設定した値よりも大きくなれば、TCP制御によってデータ出力が制限されていると判断し、該セッションがより多くの帯域でデータ出力が可能となるように、該セッションのTCP制御パラメータの値を変更する。

[0080] 但し、TCP制御パラメータを変更しても該セッションのデータ出力帯域が増加しない場合、あるいは再送タイムアウトが一定頻度以上で発生する場合には、ネットワークの輻輳を防止するため、TCPパラメータの変更を停止する。

[0081] また、該セッションの送信可能バイト数が予め設定した別の値よりも小さくなれば、TCP制御によるデータ出力をパケットスケジューラ16の割り当てウエイトによる帯域まで減少させるため、該セッションのTCP制御パラメータの値を変更する。

[0082] 前者の場合には、非輻輳時の輻輳ウインドを上げ幅を大きくしたり、輻輳時の輻輳ウインド下げ率を小さくしたりするが、後者の場合にはこの前者の場合とは逆の処理を行う。

[0083] また、本実施形態によるセッション中継装置では、パケットスケジューラ16内の割り当てウエイト変更部161が現在送信可能な帯域にしたがって動的に割り当てウエイトを変更させている。すなわち、割り当てウエイト変更部161では該セッションの送信可能バイト数が割り当てウエイトよりも大きい予め設定した値よりも大きくなれば、割り当

てウエイトで設定した帯域でのデータ出力が不可能であると判断し、割り当てウエイトを一時的に減少させる。

[0084] その後、割り当てウエイト変更部161は送信可能バイト数が予め設定した別の値よりも小さくなれば、割り当てウエイトを元の値まで順次増加させる。また、割り当てウエイト変更部161はセッション中継部12−1ー12−Nから通知される該セッションの輻輳ウインド及び往復伝播遅延時間計測値を基に該セッションが送信可能な帯域を計算する。

[0085] さらに、割り当てウエイト変更部161はこの計算値と割り当てウエイトによる帯域設定値がある一定閾値以上異なれば、割り当てウエイトによる帯域設定値が上記の計算値となるように割り当てウエイトを一時的に変更する。

(第3の実施形態)

図9は本発明の第3の実施形態によるセッション中継装置の構成を示すブロック図である。図9において、本発明の第3の実施形態によるセッション中継装置は、セッション中継部15−1ー15−Nに受信レート制御部151−1ー151−N(受信レート制御部151−2ー151−Nは図示せず)を設けた以外は図2に示す本発明の第1の実施形態によるセッション中継装置と同様の構成となっており、同一構成要素には同一符号を付してある。また、同一構成要素の動作は本発明の第1の実施形態と同様である。尚、受信レート制御部151−1ー151−Nは送信端末3からの受信レートを制御する。

[0086] この図9を参照して本発明の第3の実施形態によるセッション中継装置の動作について説明するが、ここでは、上述した本発明の第1の実施形態との相違点のみを説明する。

[0087] 本実施形態によるセッション中継装置では、送信バッファ122−1ー122−Nの空き容量がなくなると、受信セッション処理部123−1ー123−Nが送信端末3に対して広告ウインドサイズが0であることを通知し、これに応じて送信端末3はデータの送信を停止する。

[0088] 本発明の第1の実施形態では、その後、送信バッファ122−1ー122−Nに空き容量ができたとしても、送信端末3がウインド検査のためのパケットを出力してくるまでは、送信再開のためのACKを出力することができない。

[0089] これに対して、本実施形態では、パケットスケジューラ13がパケット出力の指示を行った際、受信レート制御部151−1ではパケット出力後の送信バッファ122−1ー122−Nの空き容量を検査し、もしこれが1あるいはMSS以上であれば、送信端末3に対してすばやい送信再開を促すためにACKパケットを生成する。

[0090] また、受信レート制御部151−1ー151−Nでは送信バッファ122−1ー122−Nの空き容量、あるいはその平均値が予め設定したある値以上になれば、送信バッファ122−1ー122−Nの空き容量がなくなることを防ぐために、送信端末3に対して送信帯域の低下を指示する。

[0091] これは、例えば、送信側に返送するACKパケットを重複ACKとすることや、受信パケットを廃棄すること、ACKパケットにECN(Explicit Congestion Notification)ビットを設定すること、ACKパケットを遅延させること、ACKパケットの広告ウインドを一時的に小さく書換えること等で可能である。

(第4の実施形態)

図10は本発明の第4の実施形態によるセッション中継装置の構成を示すブロック図である。図10において、本発明の第4の実施形態によるセッション中継装置は、送信iSCSI(internet Small Computer System Interface)制御部171−1ー171−N(送信iSCSI制御部171−2ー171−Nは図示せず)及び受信iSCSI制御部172−1ー172−N(受信iSCSI制御部172−2ー172−Nは図示せず)をセッション中継部17−1ー17−Nに設けた以外は上述した本発明の第3の実施形態によるセッション中継装置と同様の構成となっており、同一構成要素には同一符号を付してある。また、同一構成要素の動作は本発明の第3の実施形態によるセッション中継装置と同様である。

[0092] 送信iSCSI制御部171−1ー171−Nは受信端末2への送信レートに対してiSCSIレイヤの輻輳制御情報を反映させる。受信iSCSI制御部172−1ー172−Nは送信端末3からの受信レートに対してiSCSIレイヤの輻輳制御情報を反映させる。

[0093] この図10を参照して本発明の第4の実施形態によるセッション中継装置の動作について説明する。ここでは、本発明の第4の実施形態における本発明の第3の実施形態との相違点のみを説明する。

[0094] iSCSIレイヤにおいては、受信端末2から送信端末3に対して、受信端末2が受信可能なデータ量を通知するR2T(Ready-To-Transfer)パケットを送信することで、送信端末3と受信端末2との間で転送制御を行う。

[0095] そこで、本実施形態では、送信iSCSI制御部171-1～171-Nが受信端末2からR2Tパケットを受信すると、このR2Tパケットを受信端末2に送らず、本実施形態によるセッション中継装置にてR2Tパケット受信時に出力を行ったデータの最後のシーケンス番号を記憶しておく。

[0096] 送信iSCSI制御部171-1～171-Nは前回記憶しておいたシーケンス番号に、今回受信したR2Tパケットの送信可能データ量を加えたものを、iSCSIレイヤで送信可能なデータ量としてパケットスケジューラ13に通知する。

[0097] パケットスケジューラ13では送信待ちバイト数を、TCPレイヤが送信許可したデータ量とiSCSIレイヤが許可したデータ量との最小値とする。すなわち、図11に示すように、送信待ちバイト数＝min(連続して受信しているデータの最後尾のシーケンス番号＋1，受信端末が示した広告ウインド，該セッションの輻輳ウインド，前回R2Tパケットを受信した際のシーケンス番号＋今回受信したR2Tパケットの送信可能データ量)－送信済みシーケンス番号となる。

[0098] 受信iSCSI制御部172-1～172-Nではパケットスケジューラ13がパケット出力の指示を行った際に、パケット出力後の送信バッファ122-1～122-Nの空き容量を検査し、もしこれが予め定められ一定値以上であれば、送信端末3に対して送信バッファの空き容量をR2Tパケットとして送信する。

[0099] 上述したように、本実施形態では送信端末3と受信端末2との間でiSCSIレイヤの輻輳制御を行うのではなく、セッション中継装置でiSCSIレイヤの輻輳制御を中継することで、iSCSIレイヤにおいてもTCPセッションを中継するのと同様のスループットの向上を図ることができる。

(第5の実施形態)

図12は本発明の第5の実施形態によるセッション中継装置の構成を示すブロック図である。図12において、本発明の第5の実施形態によるセッション中継装置はセッション識別部11と、セッション送信部41-1～41-Nと、セッション受信部42-1～42-

Nと、パケットスケジューラ13と、出力制御部14とから構成されている。

[0100] セッション識別部11は到着したパケットが属するセッションを決定し、セッション送信部41−1ー41−Nは受信端末へのセッションのデータ送信処理を行い、セッション受信部42−1ー42−Nは受信端末へのセッションからのデータ受信処理を行う。

[0101] パケットスケジューラ13は各セッション送信部41−1ー41−Nからのパケット出力を制御する。出力制御部14はパケットスケジューラ13からの指示に基づいて各セッション送信部41−1ー41−Nからのパケット出力を行う。

[0102] また、セッション送信部41−1ー41−Nは送信セッション処理部411−1ー411−N(送信セッション処理部411−2ー411−Nは図示せず)と、送信データ生成部412−1ー412−N(送信データ生成部412−2ー412−Nは図示せず)と、送信バッファ413−1ー413−N(送信バッファ413−1ー413−Nは図示せず)とから構成されている。

[0103] 送信データ生成部413−1ー413−Nはアプリケーションプログラムからの送信データを送信バッファ413−1ー413−Nへと格納する。送信バッファ5−6−1は送信するデータを一時蓄えておく。送信セッション処理部411−1ー411−Nは受信端末へとデータを送信するセッションの処理を行う。

[0104] セッション受信部42−1ー42−Nは受信セッション処理部421−1ー421−N(受信セッション処理部421−2ー421−Nは図示せず)と、受信バッファ422−1ー422−N(受信バッファ422−2ー422−Nは図示せず)と、受信データ処理部423−1ー423−Nとから構成されている。

[0105] 受信セッション処理部421−1ー421−Nは受信端末からのデータの受信処理を行い、受信バッファ422−1ー422−Nは受信したデータを一時蓄えておく。受信データ処理部423−1ー423−Nは受信バッファ5−14−1からアプリケーションへ受信データを受け渡す。

[0106] TCPセッションでは、通常、送信端末と受信端末との間の双方向の通信を行うため、本実施形態では、一組の送信端末及び受信端末に対してそれぞれ1つのセッション送信部41−1ー41−N及びセッション受信部42−1ー42−Nを使用する。本実施形態においては、セッション中継装置が送信端末あるいは受信端末を兼ねている。

[0107] 図13は本発明の第5の実施形態によるセッション中継装置(送信端末)及びセッシ

ョン中継装置(受信端末)の間のデータの流れを示すブロック図である。図13において、セッション中継装置(送信端末)4−2からセッション中継装置(受信端末)4−1へデータを送る場合、セッション中継装置(送信端末)4−2のセッション送信部41−1−2から出力されたデータパケットはセッション中継装置(受信端末)4−1のセッション受信部42−1−1で受信処理が行われる。その結果、生成されたACKパケットはセッション中継装置(送信端末)4−2のセッション送信部41−1−2へと返送される。

[0108] また、セッション中継装置(受信端末)4−1からセッション中継装置(送信端末)4−2へデータを送る場合、セッション中継装置(受信端末)4−1のセッション送信部41−1−1から出力されたデータパケットはセッション中継装置(送信端末)4−2のセッション受信部42−1−2で受信処理が行われる。その結果、生成されたACKパケットはセッション中継装置(受信端末)4−1のセッション送信部41−1−1へと返送される。

[0109] 次に、図12を参照して本発明の第5の実施形態の動作について説明する。まず、送信端末から受信端末へのデータ転送に関して説明する。

[0110] アプリケーションプログラムが出力した送信データは送信データ生成部412−1ー412−Nによって送信バッファ413−1ー413−Nへと書込まれる。送信セッション処理部411−1ー411−Nは送信バッファ413−1ー413−Nから受信端末へのデータ送信処理を行う。この処理は上述した本発明の第1の実施形態と同様であるため、その説明を省略する。

[0111] ここで、パケットスケジューラ13は送信待ちバイト数を計算する際に、連続して受信しているデータの最後尾のシーケンス番号の代わりに、アプリケーションプログラムから受取ったデータの最後のシーケンス番号を用いる。

[0112] 次に、受信端末から送信端末へのデータ転送に関して説明する。受信セッション処理部421−1ー42−1Nでは受信端末から送信されたデータの受信処理を行い、正しく受取ることができたデータを受信バッファ422−1ー422−Nに格納する。

[0113] この受信処理は格納するバッファが送信バッファではなく受信バッファである点を除いて、上述した本発明の第1の実施形態と同様であるため、その説明を省略する。受信バッファ422−1ー422−Nに書込まれたデータは、受信データ処理部423−1ー423−Nによって取出されてアプリケーションプログラムへと渡される。

[0114] このように、本発明は、複数のレイヤにおける輻輳制御処理に関して、輻輳制御情報の作成のみをそれぞれのレイヤで行い、パケット出力制御処理をIPレイヤのスケジューラに集約することで、セッション中継処理の高速化を実現することができる。

[0115] また、本発明は、複数のレイヤの受信バッファ及び送信バッファをIPレイヤの送信バッファに集約することによって、バッファ間でのデータ移動をなくし、セッション中継処理の高速化を実現することができる。

請求の範囲

[1] 複数のレイヤにおける輻輳制御処理とパケット出力制御処理とを含むセッション中継処理を行うセッション中継装置において、

前記複数のレイヤ各々が前記輻輳制御情報の作成のみを行い、前記パケット出力制御処理をIP(Internet Protocol)レイヤのスケジューラに集約することを特徴とするセッション中継装置。

[2] 前記複数のレイヤに対応する受信バッファ及び送信バッファが前記IPレイヤに対応する送信バッファに集約されている請求項1記載のセッション中継装置。

[3] 送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記受信端末と前記受信端末との間の通信を実現するセッション中継装置において、

前記送信端末に向けたセッションからのデータを受信する受信セッション処理手段と、

前記受信端末に向けたセッションへとデータを送信する送信セッション処理手段と、

前記送信端末へと出力するデータを一時蓄えておく送信バッファと、

前記送信バッファからのパケット出力を制御するパケットスケジューラと、

前記パケットスケジューラの制御に応答して前記送信バッファに蓄えられたデータを出力制御する出力制御手段とを有し、

前記送信セッション処理手段において当該レイヤで出力が許可されているデータ量を計算し、これに基づいて前記パケットスケジューラが前記パケット出力を制御することを特徴とするセッション中継装置。

[4] 前記受信セッション処理手段は、TCP(Transmission Control Protocol)セッションからのデータ受信処理を行い、

前記送信セッション処理手段は、前記TCPセッションへのデータ出力処理を行い、TCPウインドフロー制御によって決定される出力可能なデータ量を前記パケットスケジューラに通知し、

前記パケットスケジューラは、通知されたデータ量を基にスケジューリング処理を行

う請求項3記載のセッション中継装置。

[5] 前記パケットスケジューラは、前記セッションに割り当てられた帯域及び帯域比率を少なくとも含む通信資源割り当て方針と、前記送信セッション処理手段から通知された送信可能データ量と、前記送信バッファ内に蓄えられているデータ量とを基にパケット出力を行うセッションを決定し、前記セッション各々からのデータ出力を制御する請求項3記載のセッション中継装置。

[6] 前記パケットスケジューラは、前記セッション各々における未使用の通信資源を蓄えておく蓄積手段をさらに備え、前記通信資源が必要になった際に前記蓄積手段に蓄えておいた通信資源を用いて通信を行う請求項3記載のセッション中継装置。

[7] 前記パケットスケジューラは、前記送信バッファ内に出力すべきデータがあり、前記出力制御手段からの出力許可データ量の制限によって未使用となった前記通信資源の帯域のみを前記蓄積手段に蓄えておくことを特徴とする請求項6記載のセッション中継装置。

[8] 前記送信セッションの制御パラメータを動的に変更する手段をさらに備え、前記パケットスケジューラからのデータ出力状況に応じて前記制御パラメータを変更する請求項3記載のセッション中継装置。

[9] 前記セッションの未使用帯域が大きくなった時に当該セッションの制御パラメータを当該セッションからの出力帯域が小さくなる方向に変更し、前記セッションの未使用帯域が小さくなった時に当該セッションの制御パラメータを当該セッションからの出力帯域が大きくなる方向に変更し、前記制御パラメータの変更によって輻輳が発生した時に前記制御パラメータの変更を停止する請求項8記載のセッション中継装置。

[10] 前記セッション各々に割り当てた帯域及び帯域比率を少なくとも含む通信資源割り当て量を動的に変更する手段をさらに備え、

前記パケットスケジューラからのデータ出力状況及び前記出力制御手段から通知される送信可能データ量に応じて前記制御パラメータを変更する請求項9記載のセッション中継装置。

[11] 前記セッションの未使用帯域が大きくなった時に当該セッションの割り当て資源を減少させ、前記セッションの未使用帯域が小さくなった時に当該セッションの割り当て

資源をその初期値を上限として増加させるとともに、前記出力制御手段から通知される送信可能データ量及びその平均のいずれかによって前記割り当て資源を増減する請求項10記載のセッション中継装置。

[12] 前記送信端末からのセッションの送信処理を制御する帯域、送信可否、送信可能データ量を少なくとも含む送信制御情報を制御する受信レート制御手段を含み、前記送信バッファの空き容量及び前記パケットスケジューラからの情報に応じて前記送信端末への送信制御情報の変更及び生成のいずれかを行う請求項3記載のセッション中継装置。

[13] 前記パケットスケジューラからのパケット出力情報を受信する手段と、パケット出力によって変更された前記送信バッファの空き容量を調べる手段とをさらに備え、パケット出力後に前記送信バッファの空き容量が一定量以上となった時に前記送信端末に対して送達確認パケットを送信して送信再開を促す請求項12記載のセッション中継装置。

[14] 前記送信バッファの空き容量及びその平均の少なくとも一方を調べる手段をさらに備え、前記空き容量に応じて前記送信端末に対して送信帯域の減少を指示する請求項12記載のセッション中継装置。

[15] 送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記送信端末と前記受信端末との間の通信を実現するセッション中継装置において、

複数のレイヤに対応して設けられかつ前記送信端末に向けたセッションからのデータを受信する受信セッション処理手段と、

前記複数のレイヤに対応して設けられかつ前記受信端末に向けたセッションへとデータを送信する送信セッション処理手段と、

前記送信端末へと出力するデータを一時蓄えておく送信バッファと、

前記送信バッファからのパケット出力を制御するパケットスケジューラとを有し、

前記送信セッション処理手段各々において当該レイヤで出力が許可されているデータ量を計算し、前記複数のレイヤ全てで共通に許可されるデータ量に基づいて前記パケットスケジューラが前記パケット出力を制御することを特徴とするセッション中継

装置。

[16] 前記レイヤとして輻輳制御を行うレイヤのひとつとしてiSCSI(internet Small Computer System Interface)レイヤを含み、当該iSCSIレイヤにおいて前記受信端末から受信する受信可能データ量を基に送信可能データ量を決定する請求項15記載のセッション中継装置。

[17] 前記パケットスケジューラからのパケット出力情報を受信する手段と、

パケット出力によって変更された前記送信バッファの空き容量を調べる手段とをさらに備え、

パケット出力後に前記送信バッファの空き容量が一定量以上となった時に前記送信端末に対して前記受信可能データ量を生成して送信再開を促す請求項15記載のセッション中継装置。

[18] 前記受信セッション処理手段は、受信したパケットを前記送信バッファに直接格納し、前記送信バッファから直接出力する請求項3記載のセッション中継装置。

[19] アプリケーションプログラムから前記送信バッファへとデータの書込みを行い、受信したデータを前記アプリケーションプログラムへと渡す請求項3記載のセッション中継装置。

[20] 複数のレイヤにおける輻輳制御処理とパケット出力制御処理とを含むセッション中継処理を行うセッション中継装置のセッション中継方法において、

前記複数のレイヤ各々で前記輻輳制御情報の作成のみを行い、前記パケット出力制御処理をIP(Internet Protocol)レイヤのスケジューラに集約することを特徴とするセッション中継方法。

[21] 21. 前記複数のレイヤに対応する受信バッファ及び送信バッファを前記IPレイヤに対応する送信バッファに集約する請求項20記載のセッション中継方法。

[22] 送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記受信端末と前記受信端末との間の通信を実現するセッション中継装置のセッション中継方法において、

前記セッション中継装置側に、

前記送信端末に向けたセッションからのデータを受信する受信セッションステップと

、

前記受信端末に向けたセッションへとデータを送信する送信セッションステップと、

前記送信端末へと出力するデータを送信バッファに一時蓄えておくステップと、

前記送信バッファからのパケット出力をパケットスケジューラにて制御するステップと

、

前記パケットスケジューラの制御に応答して前記送信バッファに蓄えられたデータを出力制御手段にて出力制御するステップとを有し、

前記送信セッション処理において当該レイヤで出力が許可されているデータ量を計算し、これに基づいて前記パケットスケジューラが前記パケット出力を制御することを特徴とするセッション中継方法。

[23] 前記受信セッションステップは、TCP(Transmission Control Protocol)セッションからのデータ受信処理を行うステップを備え、

前記送信セッションステップは、前記TCPセッションへのデータ出力処理を行い、TCPウインドフロー制御によって決定される出力可能なデータ量を前記パケットスケジューラに通知することで、前記パケットスケジューラが通知されたデータ量を基にスケジューリング処理を行うステップを備える請求項22記載のセッション中継方法。

[24] 前記パケットスケジューラが、前記セッションに割り当てられた帯域及び帯域比率を少なくとも含む通信資源割り当て方針と、前記送信セッション処理手段から通知された送信可能データ量と、前記送信バッファ内に蓄えられているデータ量とを基にパケット出力を行うセッションを決定し、前記セッション各々からのデータ出力を制御するステップをさらに備える請求項22記載のセッション中継方法。

[25] 前記パケットスケジューラには前記セッション各々における未使用の通信資源を蓄えておく蓄積手段が含まれ、

前記パケットスケジューラにおいて前記通信資源が必要になった際に前記蓄積手段に蓄えておいた通信資源を用いて通信を行うステップをさらに備えた請求項22のいずれか記載のセッション中継方法。

[26] 前記パケットスケジューラが、前記送信バッファ内に出力すべきデータがあり、前記出力制御手段からの出力許可データ量の制限によって未使用となった前記通信資

源の帯域のみを前記蓄積手段に蓄えておくステップをさらに備えた請求項25記載のセッション中継方法。

[27] 前記送信セッションの制御パラメータを動的に変更する手段が、前記パケットスケジューラからのデータ出力状況に応じて前記制御パラメータを変更するステップをさらに備えた請求項22記載のセッション中継方法。

[28] 前記セッションの未使用帯域が大きくなった時に当該セッションの制御パラメータを当該セッションからの出力帯域が小さくなる方向に変更し、前記セッションの未使用帯域が小さくなった時に当該セッションの制御パラメータを当該セッションからの出力帯域が大きくなる方向に変更し、前記制御パラメータの変更によって輻輳が発生した時に前記制御パラメータの変更を停止するステップをさらに備えた請求項27記載のセッション中継方法。

[29] 前記セッション各々に割り当てた帯域及び帯域比率を少なくとも含む通信資源割り当て量を動的に変更する手段が、前記パケットスケジューラからのデータ出力状況及び前記出力制御手段から通知される送信可能データ量に応じて前記制御パラメータを変更するステップをさらに備えた請求項28記載のセッション中継方法。

[30] 前記セッションの未使用帯域が大きくなった時に当該セッションの割り当て資源を減少させ、前記セッションの未使用帯域が小さくなった時に当該セッションの割り当て資源をその初期値を上限として増加させるとともに、前記出力制御手段から通知される送信可能データ量及びその平均のいずれかによって前記割り当て資源を増減するステップをさらに備えた請求項29記載のセッション中継方法。

[31] 前記送信端末からのセッションの送信処理を制御する帯域、送信可否、送信可能データ量を少なくとも含む送信制御情報を制御する受信レート制御手段が、前記送信バッファの空き容量及び前記パケットスケジューラからの情報に応じて前記送信端末への送信制御情報の変更及び生成のいずれかを行うステップをさらに備えた請求項22から請求項29のいずれか記載のセッション中継方法。

[32] パケット出力後に前記送信バッファの空き容量が一定量以上となった時に前記送信端末に対して送達確認パケットを送信して送信再開を促すステップをさらに備えた請求項31記載のセッション中継方法。

[33] 前記送信バッファの空き容量及びその平均の少なくとも一方を調べる手段で調べられた前記空き容量に応じて前記送信端末に対して送信帯域の減少を指示するステップをさらに備えた請求項31記載のセッション中継方法。

[34] 送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記送信端末と前記受信端末との間の通信を実現するセッション中継装置のセッション中継方法において、

前記セッション中継装置側に、

複数のレイヤ各々において前記送信端末に向けたセッションからのデータを受信する受信セッションステップと、

前記複数のレイヤ各々において前記受信端末に向けたセッションへとデータを送信する送信セッションステップと、

前記送信端末へと出力するデータを送信バッファに一時蓄えておくステップと、

前記送信バッファからのパケット出力をパケットスケジューラにて制御するステップとを備え、

前記送信セッション処理各々において当該レイヤで出力が許可されているデータ量を計算し、前記複数のレイヤ全てで共通に許可されるデータ量に基づいて前記パケットスケジューラが前記パケット出力を制御することを特徴とするセッション中継方法。

[35] 前記レイヤとして輻輳制御を行うレイヤのひとつとしてiSCSI(internet Small Computer System Interface)レイヤを含み、

当該iSCSIレイヤにおいて前記受信端末から受信する受信可能データ量を基に送信可能データ量を決定するステップをさらに備えた請求項34記載のセッション中継方法。

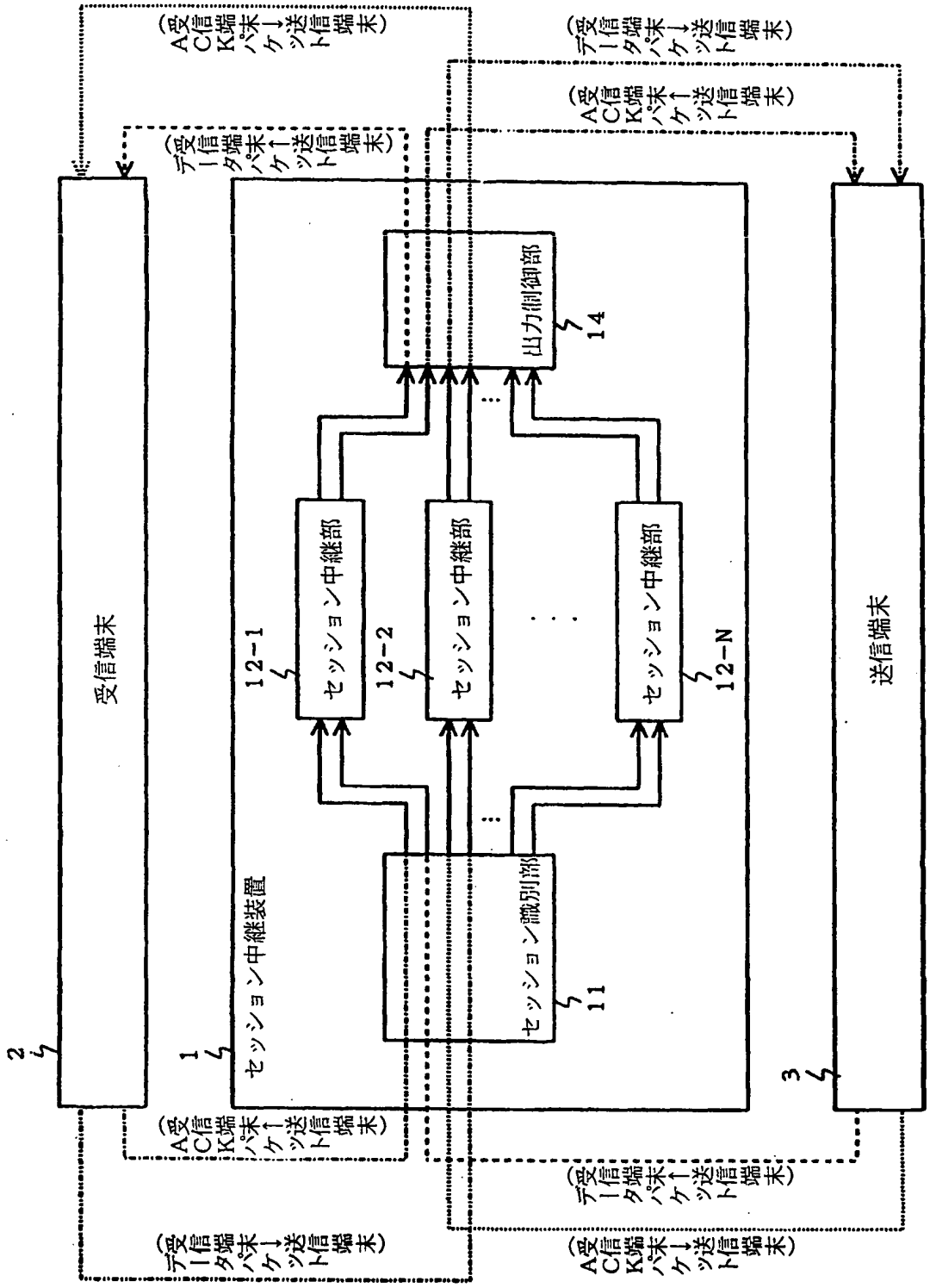
[36] パケット出力後に前記送信バッファの空き容量が一定量以上となった時に前記送信端末に対して前記受信可能データ量を生成して送信再開を促すステップをさらに備えた請求項34記載のセッション中継方法。

[37] 前記受信セッションステップは、受信したパケットを前記送信バッファに直接格納し、前記送信バッファから直接出力するステップをさらに備えた請求項22記載のセッシ

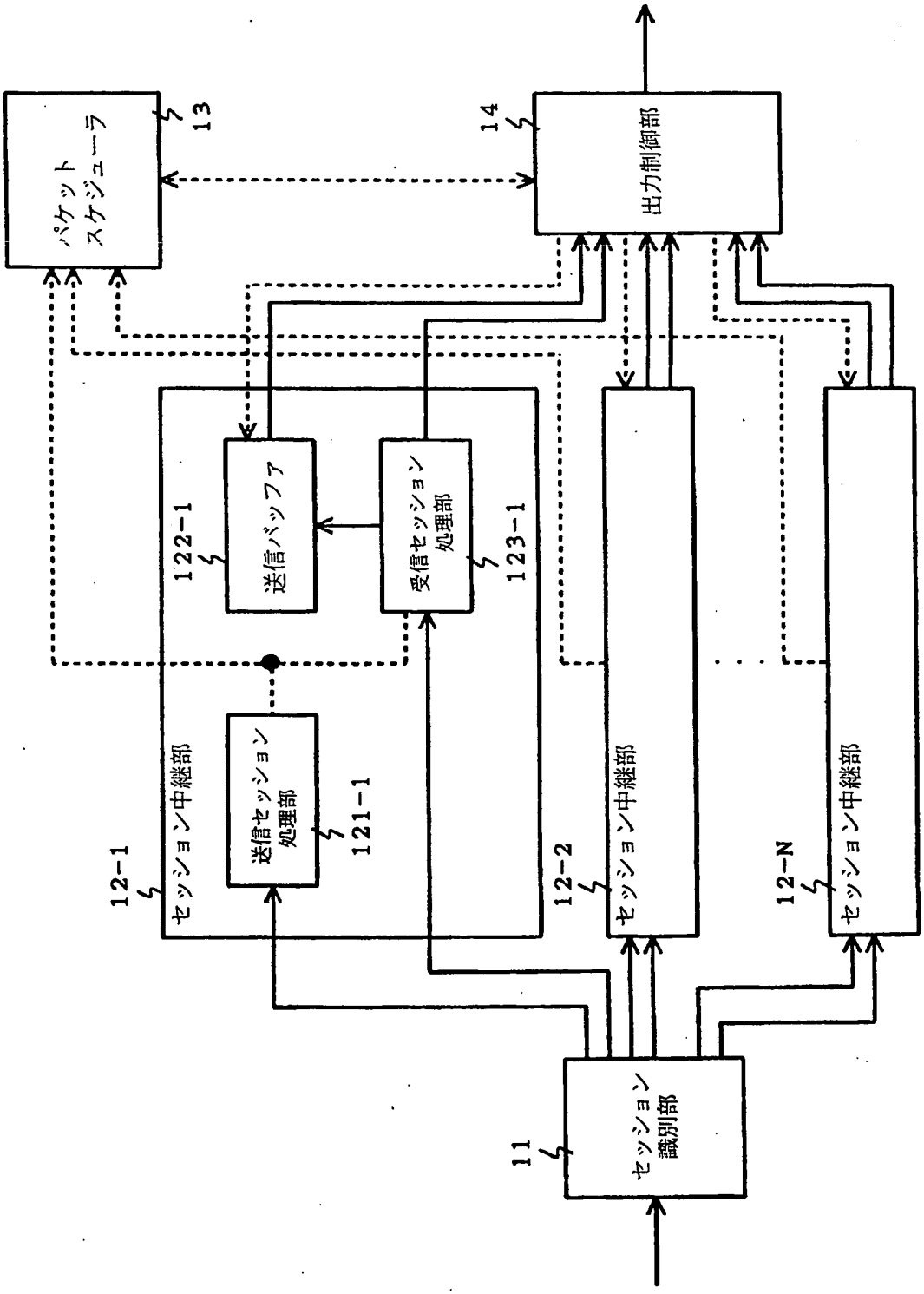
ョン中継方法。

[38] アプリケーションプログラムから前記送信バッファへとデータの書込みを行い、受信したデータを前記アプリケーションプログラムへと渡すステップをさらに備えた請求項22記載のセッション中継方法。

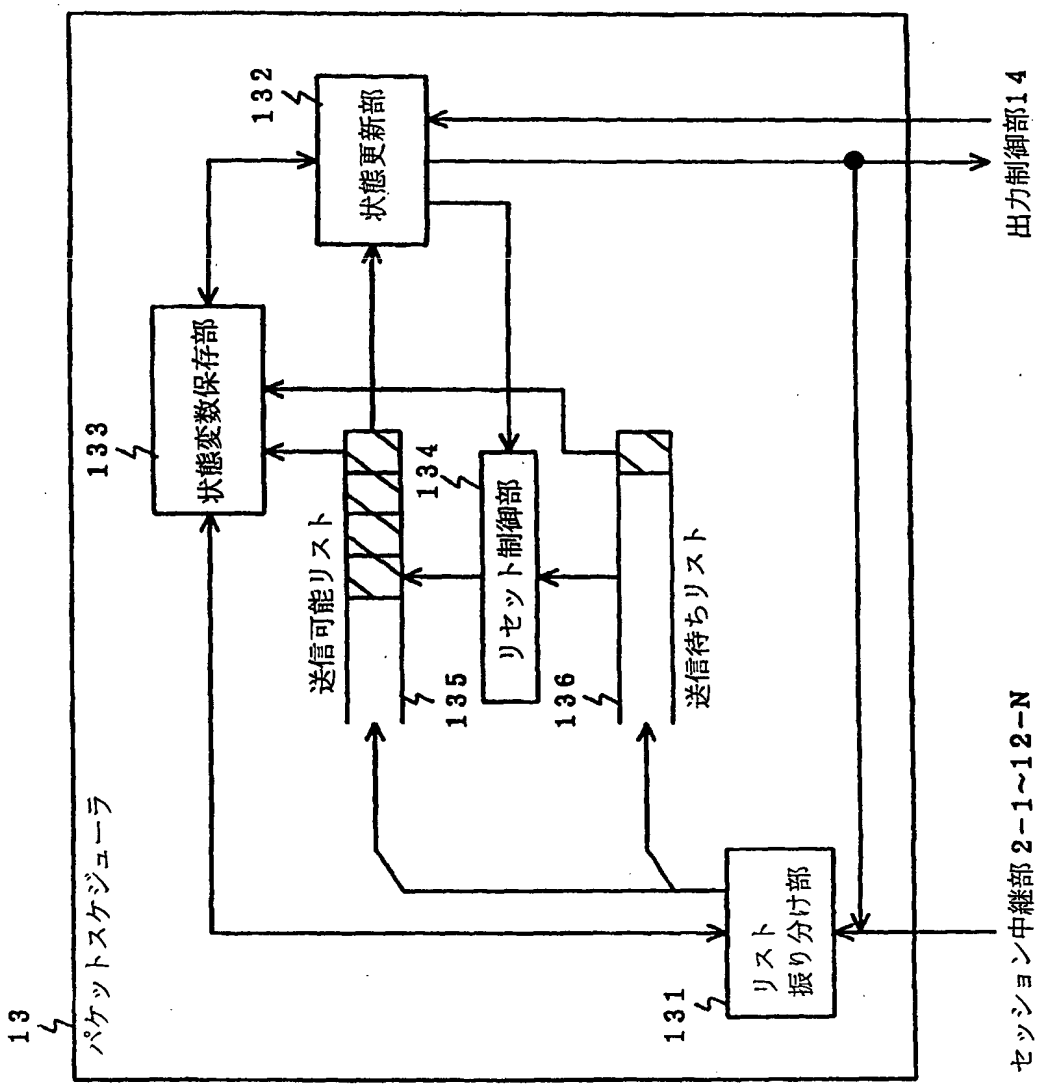
[図1]


受信端末
2
セッション中継装置
1
セッション識別部
11
出力側御部
14
セッション中継部
12-1
セッション中継部
12-2
セッション中継部
12-N
送信端末
3
(受信データ端末→送信端末)
(受信A/C/K端末→送信端末)
(受信データ端末→送信端末)
(受信A/C/K端末→送信端末)
(受信A/C/K端末→送信端末)
(受信データ端末→送信端末)
(受信A/C/K端末→送信端末)
(受信データ端末→送信端末)

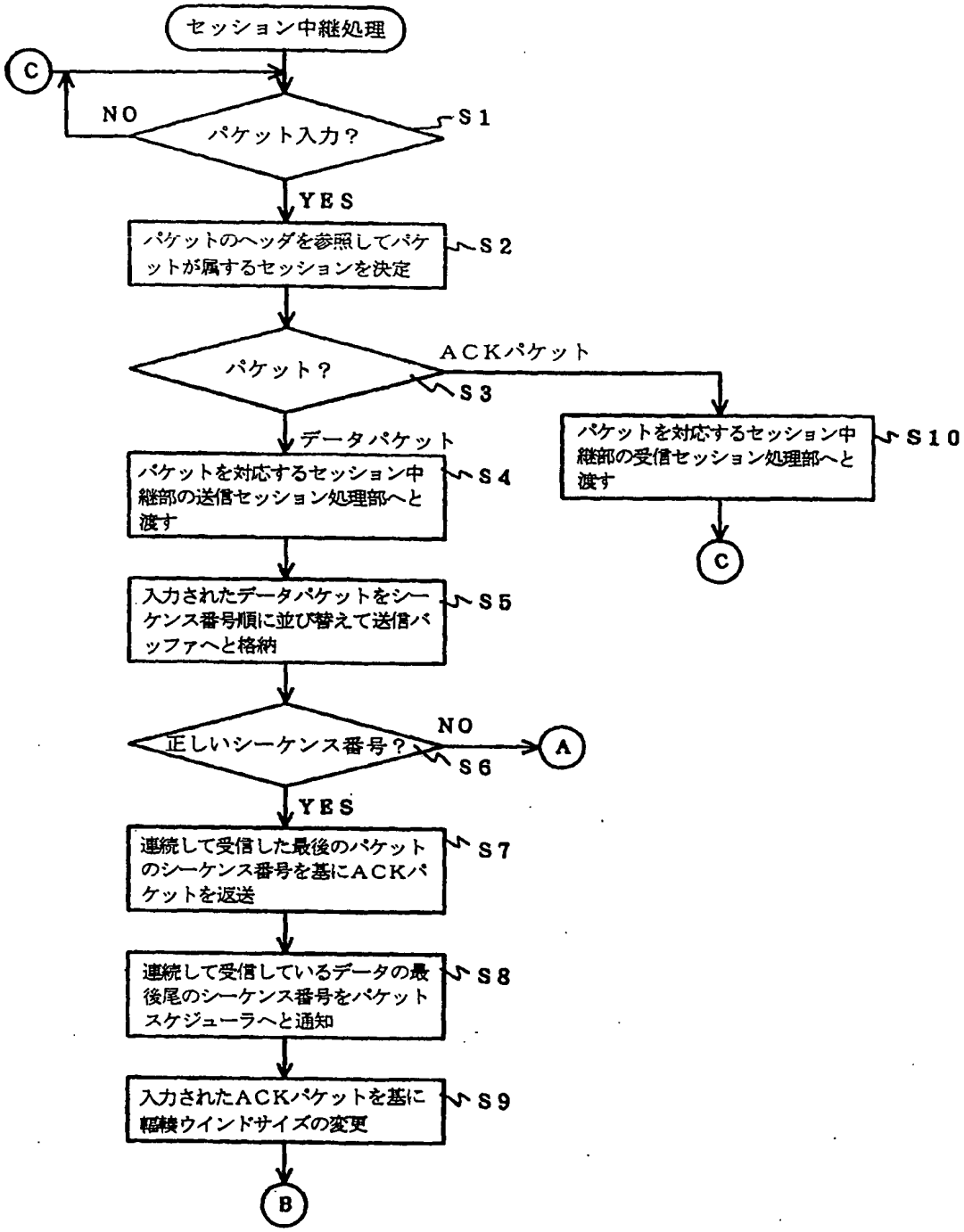
[図2]


パケット
スケジューラ
13
出力制御部
14
送信バッファ
122-1
受信セッション
処理部
123-1
送信セッション
処理部
121-1
セッション中継部
12-1
セッション中継部
12-2
セッション中継部
12-N
セッション
識別部
11

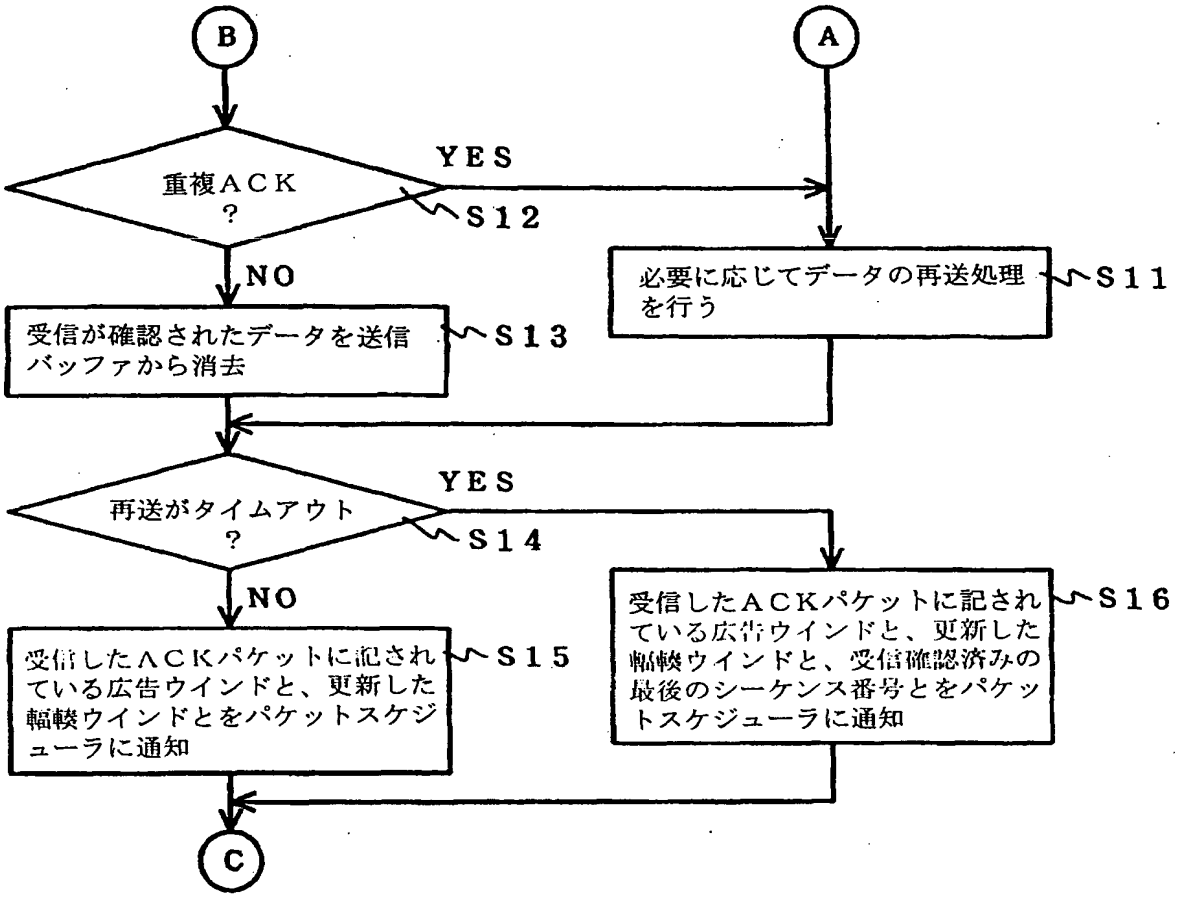
[図3]


13
パケットスケジューラ
131
リスト振り分け部
132
状態更新部
133
状態変数保存部
134
リセット制御部
135
送信可能リスト
136
送信待ちリスト
出力制御部14
セッション中継部2-1~12-N

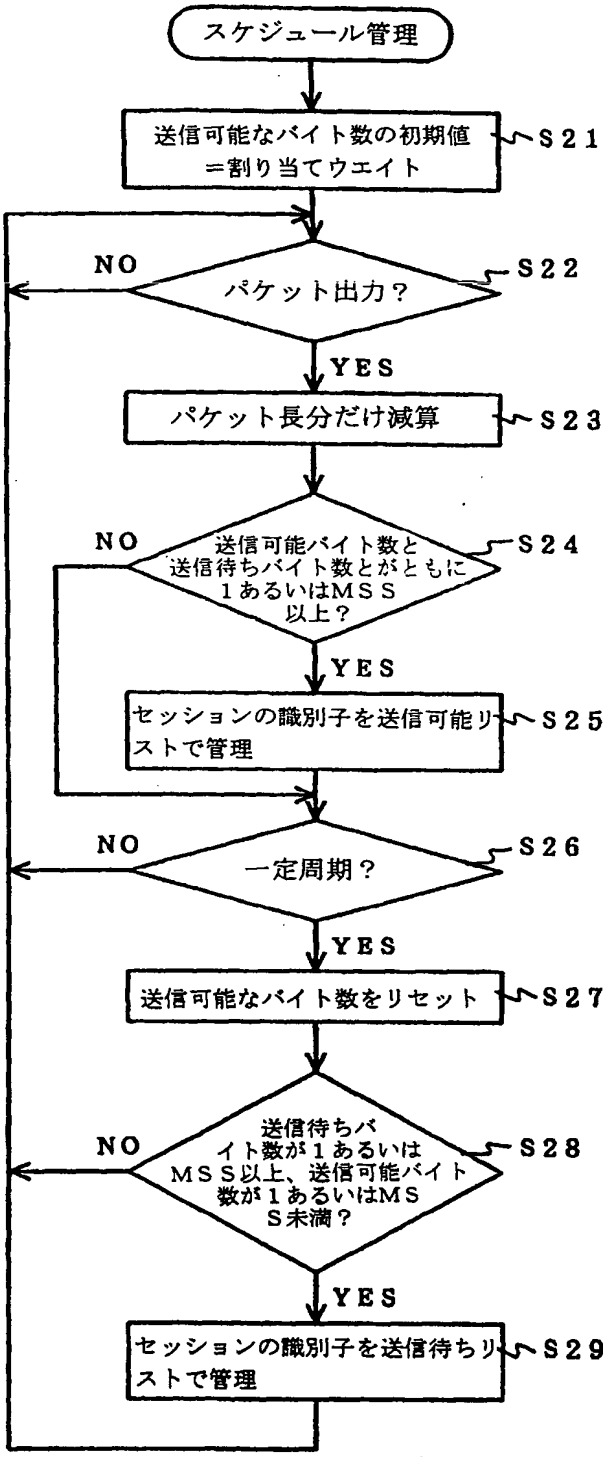
[図4]


セッション中継処理
C
NO
パケット入力？
S1
YES
パケットのヘッダを参照してパケットが属するセッションを決定
S2
パケット？
ACKパケット
S3
データパケット
パケットを対応するセッション中継部の送信セッション処理部へと渡す
S4
パケットを対応するセッション中継部の受信セッション処理部へと渡す
S10
C
入力されたデータパケットをシーケンス番号順に並び替えて送信バッファへと格納
S5
正しいシーケンス番号？
NO
A
S6
YES
連続して受信した最後のパケットのシーケンス番号を基にACKパケットを返送
S7
連続して受信しているデータの最後尾のシーケンス番号をパケットスケジューラへと通知
S8
入力されたACKパケットを基に輻輳ウインドサイズの変更
S9
B

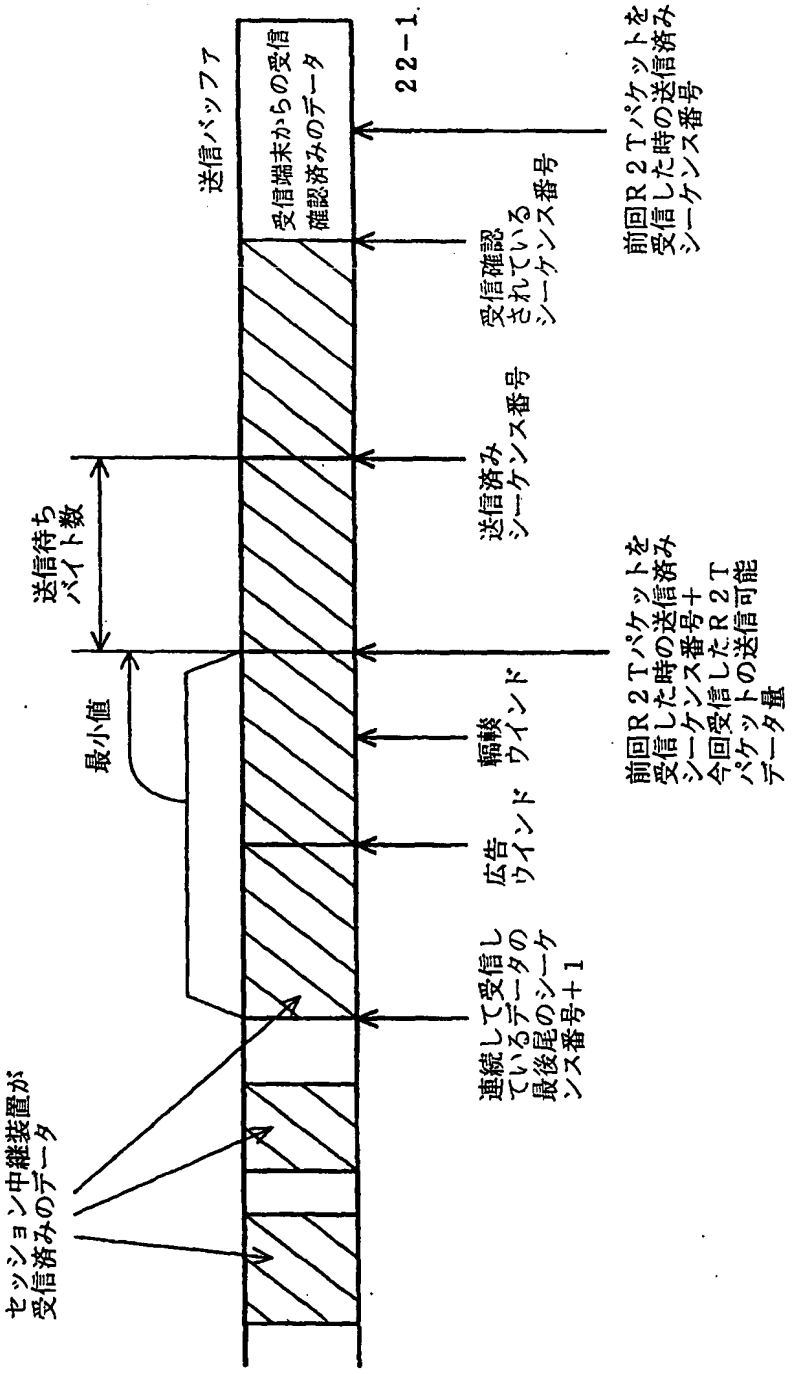
[図5]


B
A
重複ＡＣＫ
？
YES
S12
NO
必要に応じてデータの再送処理を行う
S11
受信が確認されたデータを送信バッファから消去
S13
再送がタイムアウト
？
YES
S14
NO
受信したＡＣＫパケットに記されている広告ウインドと、更新した輻輳ウインドと、受信確認済みの最後のシーケンス番号とをパケットスケジューラに通知
S16
受信したＡＣＫパケットに記されている広告ウインドと、更新した輻輳ウインドとをパケットスケジューラに通知
S15
C

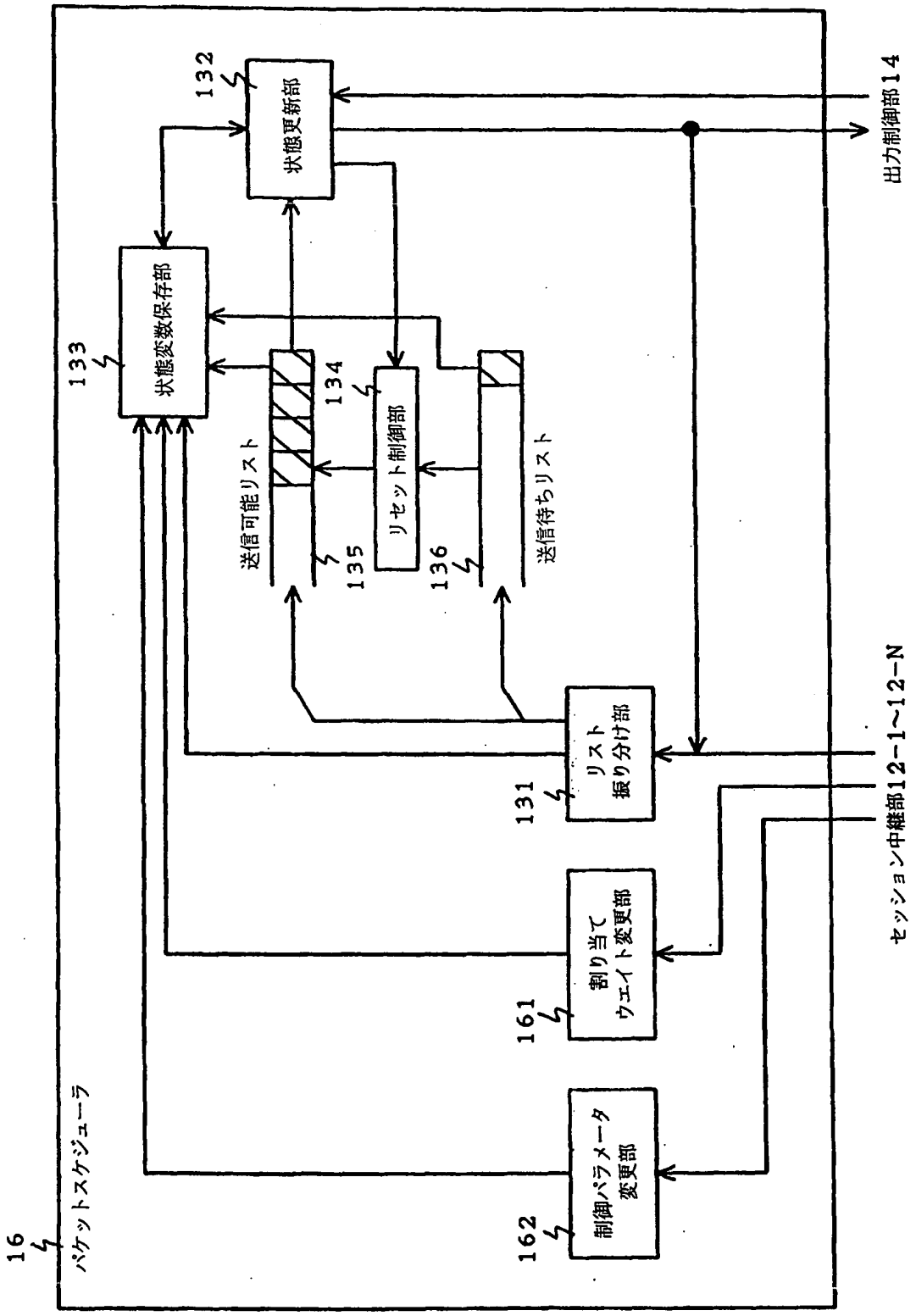
[図6]


スケジュール管理
送信可能なバイト数の初期値＝割り当てウエイト
S21
NO
パケット出力？
S22
YES
パケット長分だけ減算
S23
NO
送信可能バイト数と送信待ちバイト数とがともに１あるいはＭＳＳ以上？
S24
YES
セッションの識別子を送信可能リストで管理
S25
NO
一定周期？
S26
YES
送信可能なバイト数をリセット
S27
NO
送信待ちバイト数が１あるいはＭＳＳ以上、送信可能バイト数が１あるいはＭＳＳ未満？
S28
YES
セッションの識別子を送信待ちリストで管理
S29

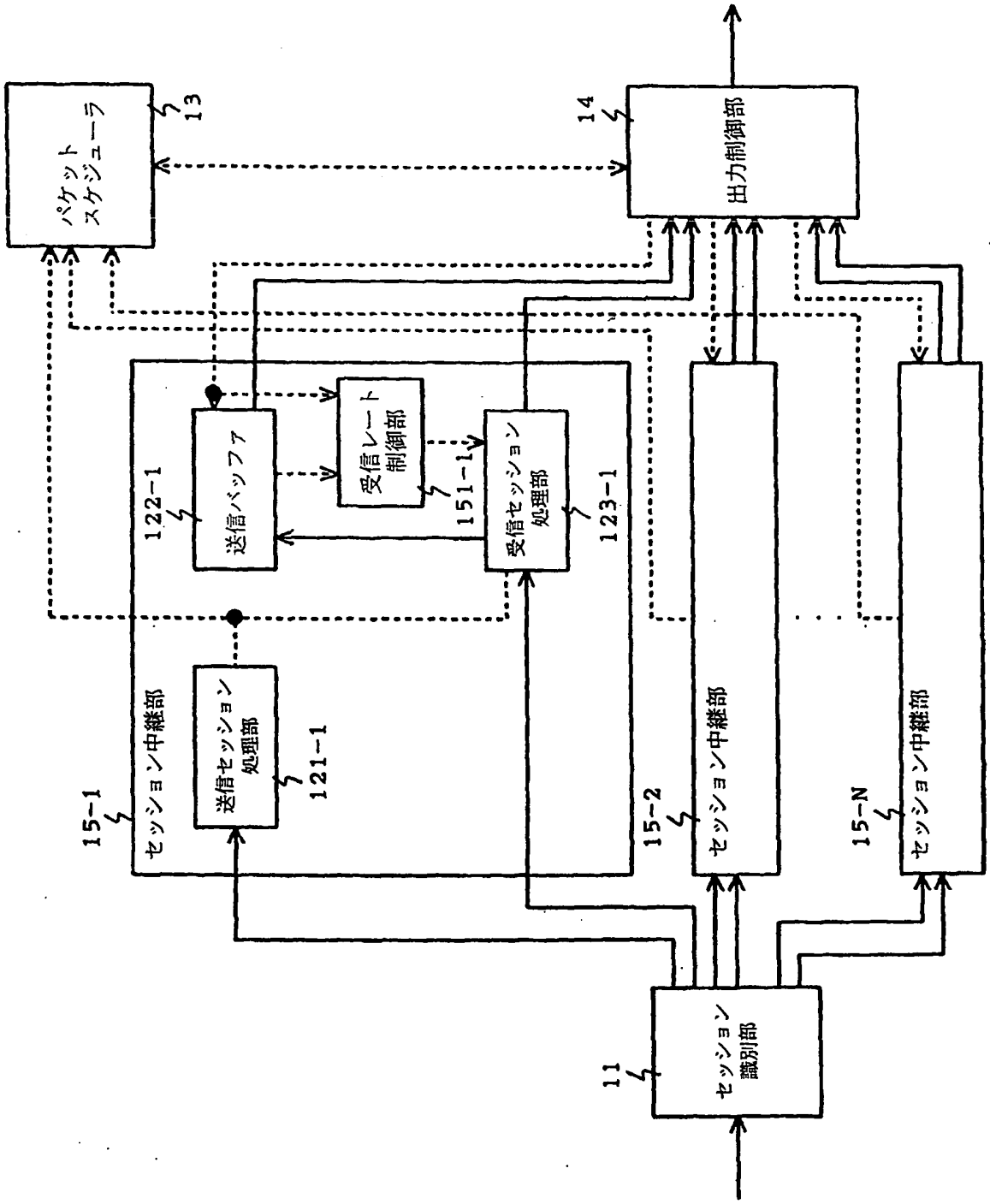
[図7]


送信バッファ
受信端末からの受信確認済みのデータ
22-1
前回R2Tパケットを受信した時の送信済みシーケンス番号
受信確認されているシーケンス番号
送信済みシーケンス番号
送信待ちバイト数
最小値
輻輳ウインド
広告ウインド
前回R2Tパケットを受信した時の送信済みシーケンス番号+今回受信したR2Tパケットの送信可能データ量
連続して受信しているデータの最後尾のシーケンス番号+1
セッション中継装置が受信済みのデータ

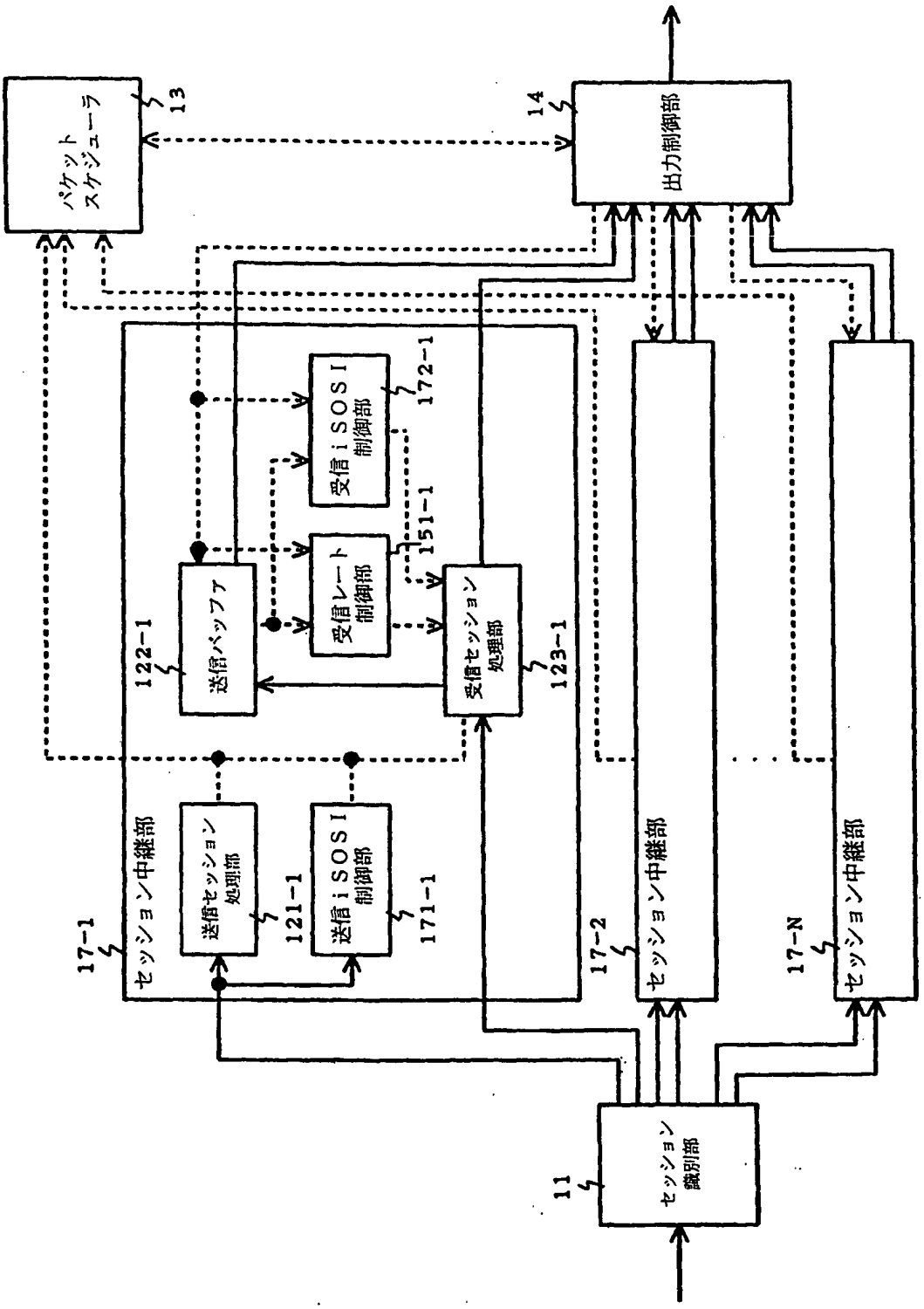
[図8]


16
パケットスケジューラ
132
状態更新部
133
状態変数保存部
134
リセット制御部
135
送信可能リスト
136
送信待ちリスト
131
リスト振り分け部
161
割り当てウェイト変更部
162
制御パラメータ変更部
出力制御部14
セッション中継部12-1~12-N

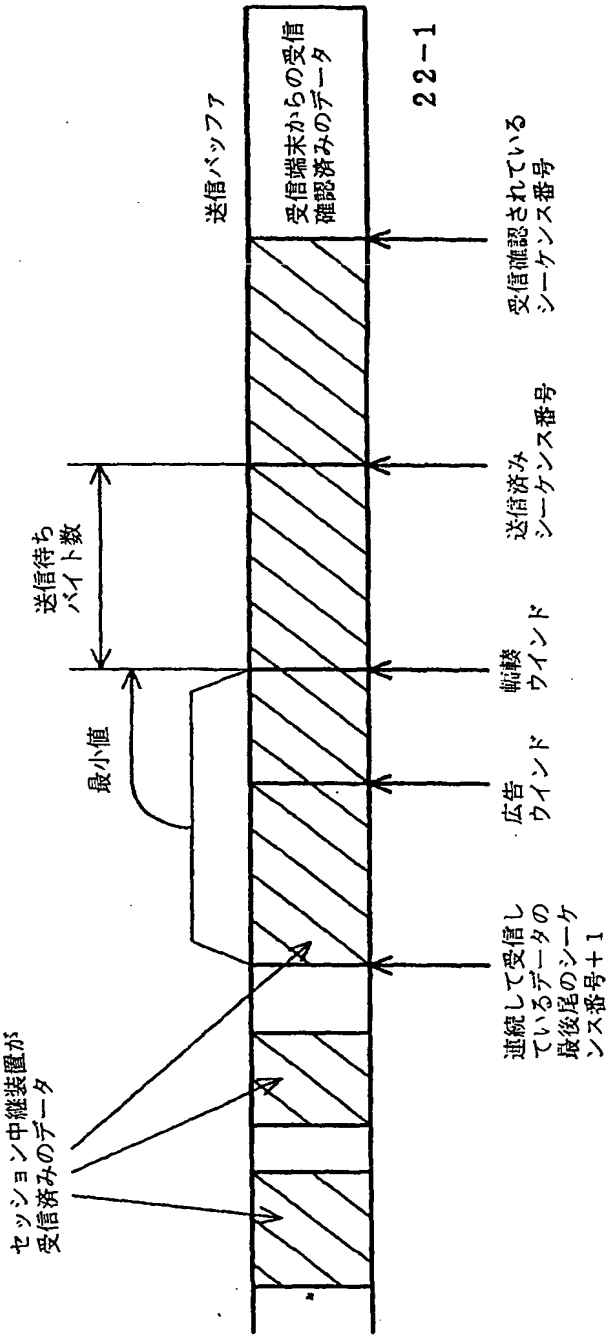
[図9]


パケット スケジューラ
13
14
出力制御部
送信バッファ
122-1
受信レート 制御部
151-1
受信セッション 処理部
123-1
送信セッション 処理部
121-1
セッション中継部
15-1
セッション中継部
15-2
セッション中継部
15-N
セッション 識別部
11

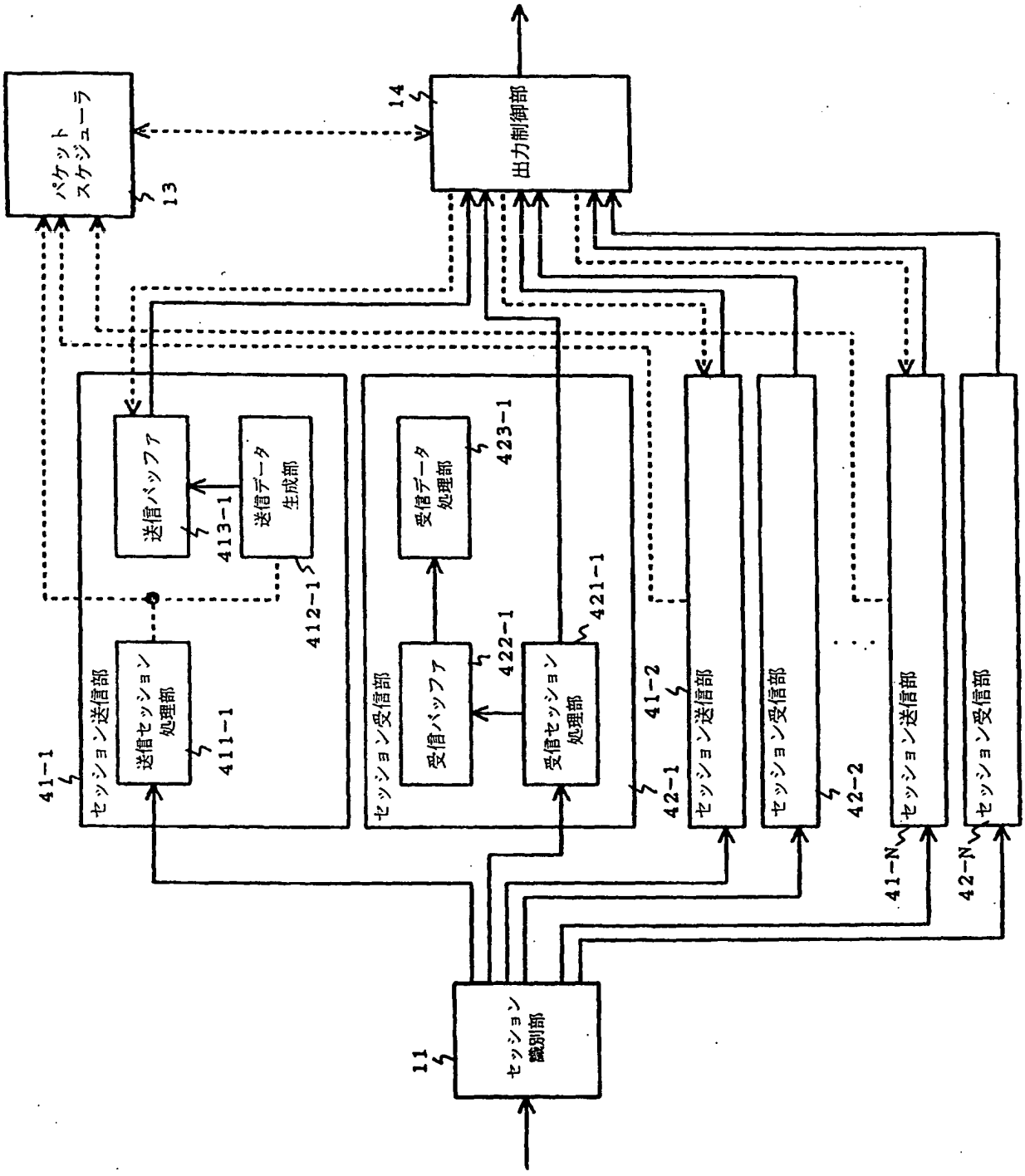
[図10]


パケット
スケジューラ
13
出力制御部
14
受信iSOSI
制御部
172-1
151-1
受信レート
制御部
送信バッファ
122-1
受信セッション
処理部
123-1
送信セッション
処理部
121-1
送信iSOSI
制御部
171-1
セッション中継部
17-1
セッション中継部
17-2
セッション中継部
17-N
セッション
識別部
11

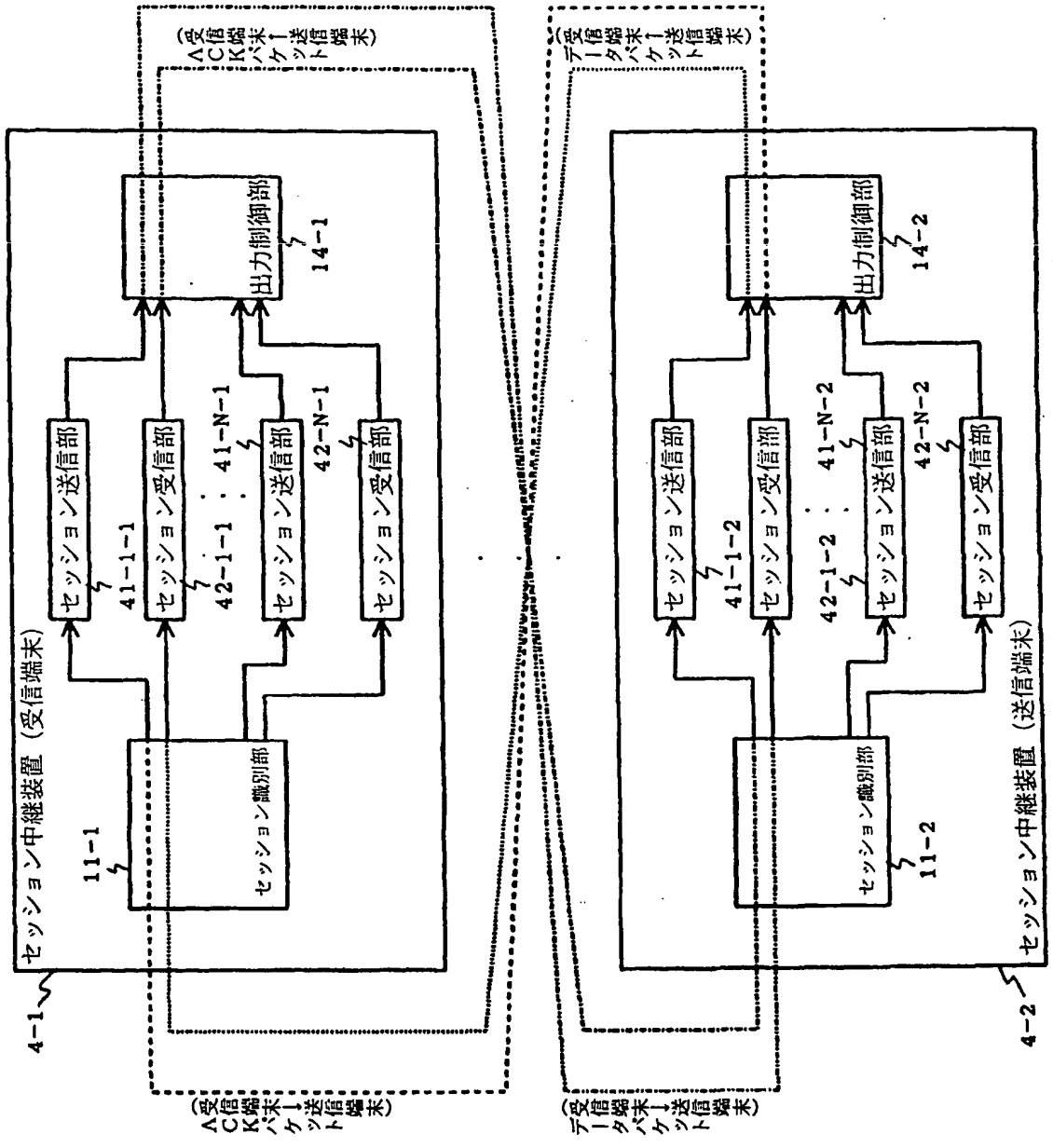
[図11]


送信バッファ
受信端末からの受信確認済みのデータ
22-1
受信確認されているシーケンス番号
送信済みシーケンス番号
輻輳ウインドウ
広告ウインドウ
連続して受信しているデータの最後尾のシーケンス番号＋1
送信待ちバイト数
最小値
セッション中継装置が受信済みのデータ

[図12]


パケット
スケジューラ
13
14
出力制御部
送信バッファ
413-1
送信データ
生成部
412-1
受信データ
処理部
423-1
421-1
422-1
受信バッファ
受信セッション
処理部
送信セッション
処理部
411-1
セッション送信部
41-1
セッション受信部
41-2
42-1
セッション送信部
セッション受信部
42-2
セッション送信部
41-N
セッション受信部
42-N
セッション
識別部
11

[図13]


セッション中継装置（受信端末）
4-1
11-1
セッション識別部
41-1-1
セッション送信部
42-1-1
セッション受信部
41-N-1
セッション送信部
42-N-1
セッション受信部
出力制御部
14-1
（ACK端末→データ端末）
（データ端末→ACK端末）
41-1-2
セッション送信部
42-1-2
セッション受信部
41-N-2
セッション送信部
42-N-2
セッション受信部
出力制御部
14-2
セッション識別部
11-2
セッション中継装置（送信端末）
4-2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/010604

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl[7] H04L12/56

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl[7] H04L12/56

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2004
Kokai Jitsuyo Shinan Koho 1971-2004 Toroku Jitsuyo Shinan Koho 1994-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y A	JP 10-233802 A (Lucent Technologies Inc.), 02 September, 1998 (02.09.98), Full text; Fig. 2 & EP 872988 A2 & CA 2227244 A & US 6092115 A	3-5,18,19, 22-24,37,38 6-8,25-27 9-14,28-33
Y	JP 2000-49787 A (Hitachi, Ltd.), 18 February, 2000 (18.02.00), Par. No. [0032]; Fig. 2 & US 6657964 B1	6,7,25,26
Y	JP 2002-344500 A (NEC Corp.), 29 November, 2002 (29.11.02), Full text; all drawings (Family: none)	8,27

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 August, 2004 (19.08.04)	07 September, 2004 (07.09.04)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/010604

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2000-49856 A (Toshiba Corp.), 18 February, 2000 (18.02.00), Claims 1 to 4; all drawings (Family: none)	1,2,15-17, 20,21,34-36
A	JP 2001-358771 A (NEC Corp.), 26 December, 2001 (26.12.01), Claim 1; all drawings & EP 971518 A2 & CA 2277229 A1 & US 6415313 B1	1,2,15-17, 20,21,34-36
A	Masayoshi KOBAYASHI et al., "Koritsuteki na Prefetch o Okonau Pro Active Cash Network Kose", Shingaku Giho IN2000-150, 22 November, 2000 (22.11.00), page 106, 4.3 to page 107, 5.2, Fig.4	1-38

国際調査報告

国際出願番号　PCT／JP2004／010604

A.　発明の属する分野の分類（国際特許分類（IPC））

Int. Cl⁷ H04L 12／56

B.　調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int. Cl⁷ H04L 12／56

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922－1996年
日本国公開実用新案公報	1971－2004年
日本国実用新案登録公報	1996－2004年
日本国登録実用新案公報	1994－2004年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C.　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
X	JP　10－233802　A（ルーセント　テクノロジーズ　インコーポレイテッド） 1998.09.02，全文，第2図　&	3-5, 18, 19, 22-24, 37, 38
Y	EP　872988　A2　& CA　2227244　A　&	6-8, 25-27
A	US　6092115　A	9-14, 28-33
Y	JP　2000－49787　A（株式会社日立製作所） 2000.02.18，【0032】，第2図　& US　6657964　B1	6, 7, 25, 26

☒ C欄の続きにも文献が列挙されている。　☐ パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日 19.08.2004	国際調査報告の発送日　07.9.2004
国際調査機関の名称及びあて先 日本国特許庁（ISA／JP） 郵便番号100－8915 東京都千代田区霞が関二丁目4番3号	特許庁審査官（権限のある職員） 石井　研一　5X　3250 電話番号　03－3581－1101　内線　3555

様式PCT／ISA／210（第2ページ）（2004年1月）

Ｃ（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
Ｙ	ＪＰ　２００２－３４４５００　Ａ（日本電気株式会社） ２００２．１１．２９，全文，全図（ファミリーなし）	8, 27
Ａ	ＪＰ　２０００－４９８５６　Ａ（株式会社東芝） ２０００．０２．１８，【請求項１】～【請求項４】，全図 （ファミリーなし）	1, 2, 15-17, 20, 21, 34-36
Ａ	ＪＰ　２００１－３５８７７１　Ａ（日本電気株式会社） ２００１．１２．２６，【請求項１】，全図　＆ ＥＰ　９７１５１８　Ａ２　＆ ＣＡ　２２７７２２９　Ａ１　＆ ＵＳ　６４１５３１３　Ｂ１	1, 2, 15-17, 20, 21, 34-36
Ａ	小林正好他，効率的なプリフェッチを行うプロアクティブキャッシュネットワーク構成，信学技報ＩＮ２０００－１５０， ２０００．１１．２２， 第１０６頁４．３～第１０７頁５．２，図４	1-38